UNIVERSIDADE FEDERAL DO PARANÁ

ZAUDIR DAL CORTIVO

ANÁLISE DISCRIMINANTE LINEAR DE FISHER COMBINADA COM ALGORITMO *K*-SEGMENTOS.

CURITIBA

2015

ZAUDIR DAL CORTIVO

ANÁLISE DISCRIMINANTE LINEAR DE FISHER COMBINADA COM ALGORITMO *K*-SEGMENTOS.

Tese de doutorado apresentada ao programa de Pós-graduação em Métodos Numéricos em Engenharia, do Setor de Exatas e do Setor de Tecnologia da Universidade Federal do Paraná, na linha de Pesquisa de Métodos Estatísticos Aplicados à Engenharia, como requisito parcial à obtenção do título de Doutor em Métodos Numéricos em Engenharia.

Orientador: Prof. Dr. Jair Mendes Marques.

CURITIBA

2015

O48 Dal Cortivo, Zaudir

Análise discriminante linear de *fisher* combinada com algoritmo *k*-segmentos – Curitiba: UFPR/ITTI, 2015.
160 : il.; tabs. : color. : 30 cm.

Tese – Universidade Federal do Paraná, Programa de Pós-graduação em Métodos Numéricos em Engenharia, setor de Exatas e setor de Tecnologia.
Orientador: Prof. Dr. Jair Mendes Marques.
Bibliografia: p.126-130.

1. Análise multivariada. 2. Análise discriminante. I. Universidade Federal do Paraná. II. Marques, Jair Mendes. III. Título.

CDD 634

TERMO DE APROVAÇÃO

ZAUDIR DAL CORTIVO

ANÁLISE DISCRIMINANTE LINEAR DE FISHER COMBINADA COM ALGORITMO K-SEGMENTOS

Tese aprovada como requisito parcial para obtenção do grau de doutor no Programa de Pós-Graduação em Métodos Numéricos em Engenharia, da Universidade Federal do Paraná, pela seguinte banca examinadora:

Prof. Dr. Jair Mendes Marques
Orientador – membro do PPGMNE/UFPR

Prof. Dr. Anselmo Chaves Neto
Membro do PPGMNE/UFPR

Prof. Dr. Inácio Andruski
Membro da UTFPR – Curitiba/PR

Prof. Dr. Maurício Koubay do Amaral
Membro da UTFPR – Curitiba/PR

Prof.ª Dr.ª Sachiko Araki Lira
Membro do PPGMNE/UFPR

Curitiba, 04 de dezembro de 2015.

AGRADECIMENTOS

Ao Deus único, o todo poderoso, Senhor dos senhores, criador dos Céus e da Terra, que sempre tem me abençoado e fortalecido.

Ao meu orientador, Prof. Dr. Jair Mendes Marques, pelo acompanhamento e orientação e ao Prof. Dr. Anselmo Chaves Neto, pelas suas contribuições.

À minha esposa Carmen e aos meus filhos Pedro e Daniel, pelo apoio e compreensão das minhas ausências.

Ao Governo do Estado do Paraná e à Secretaria de Educação do Estado do Paraná (SEED/PR) pela licença parcial para estudo.

Ao Luiz Hamilton Guedes e amigos, pelas orações em meu favor.

“Eu segurei muitas coisas em minhas mãos, e eu perdi tudo; mas tudo que eu coloquei nas mãos de Deus eu ainda possuo”.

Martin Luther King

"Porque eu estou bem certo de que nem a morte, nem a vida, nem os anjos, nem os principados, nem as coisas do presente, nem do porvir, nem os poderes, nem a altura, nem a profundidade, nem qualquer outra criatura poderá separar-nos do amor de Deus, que está em Cristo Jesus, nosso Senhor." Rm 9: 38-39.

RESUMO

A classificação de dados amostrais em categorias tem aplicações em diversas áreas da ciência e existem várias técnicas que podem ser aplicadas para esse fim, como a Análise Discriminante Linear de Fisher (FDA) e a Análise Discriminante Local de Fisher (LFDA). A eficiência de classificação dos dados é relevante para qualquer técnica. Diferentes técnicas ou apenas modificações de métodos já existentes têm sido devolvidas no meio científico, com o objetivo de oferecer métodos mais eficientes para a classificação amostral de dados. Em geral, as metodologias para análise discriminante são eficientes em algumas situações e em outras não. As técnicas FDA e LFDA podem ser menos eficientes quando a dispersão nas classes é alta ou quando os dados estão dispersos em forma linear ou não linear. Para essas situações, o uso apenas dos centroides para associar uma nova observação amostral à classe cujo centroide é mais próximo pode não ser o método mais eficaz. A proposta deste trabalho é efetuar modificação nas técnicas FDA e LFDA substituindo os centroides por uma linha poligonal gerada pelo algoritmo *k*-segmentos. Esse método consiste em ajustar para cada classe, definida *a priori*, uma linha poligonal e uma nova observação é classificada na classe cuja distância ortogonal à linha poligonal é a menor. Experimentalmente, o algoritmo é aplicado para diversos conjuntos de dados amostrais e os resultados são comparados com a taxa aparente de erro. Os resultados experimentais obtidos com aplicação da técnica *k*-segmentos, mostraram-se mais eficientes que o uso dos centroides, tanto para FDA como para LFDA. Os testes realizados apresentaram ganhos significativos na percentagem de observações amostrais classificadadas corretamente, principalmente quando os escorres tem dispersão linear ou não linear. Outra vantagem é que ele não utiliza apenas um ponto central, como os centroides, mas diversos pontos que 'passam' no meio dos escorres de cada classe, o que pode possibilitar melhor eficiência de classificação.

Palavras-chave: Análise Discriminante de Fisher. Curvas Principais. Algoritmo *k*-segmentos. Análise Discriminante Local de Fisher.

ABSTRACT

The classification of sample data has applications in several areas of science and there are various techniques that can be applied to this order, as Fisher's Linear discriminant analysis (FDA) and the Local Fisher discriminant analysis (LFDA). The efficieny of data classification is relevant to any technique. Different techniques or just modifications of existing methods, have been returned in the scientific world, with the aim of offering more efficient methods for the sampling of data classification. In General, the methodologies for discriminant analysis are effective in some situations and not others. The FDA and LFDA techniques can be less efficient when the dispersion in the classes is high or when the data are dispersed in linear or nonlinear form. For these situations the use of only the centróides to associate a new sample observation to class whose centroid is closest may not be the most efficient method. The purpose of this work is to make modification in FDA and replacing the LFDA centróides by a polygonal line generated by the algorithm *k*-segments. This method consists in adjusting to each class, defined *a priori*, a polygonal line, and a new observation is classified in class whose orthogonal distance the polygonal line is the smallest. Experimentally, the algorithm is applied to several sets of sample data and the results are compared with the apparent rate of error. The experimental results obtained with application of *k*-segments, were more efficient than the use of centroides, both for FDA and for LFDA. The tests showed significant gains in the percentage of sample observations classificadadas correctly, especially when the escores have linear or non-linear dispersion. Another advantage is that it not only uses a central point, like the centroids, but several points to ' pass ' in the middle of the scores of each class, which can enable better efficiency.

Keywords: Fisher Discriminant Analysis. Principal Curves. Algorithm *k*-segments. Local Fisher Discriminant analysis.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

AER	-	Taxa Real de Erro
APER	-	Taxa Aparente de Erro
CP	-	Curva Principal
FDA	-	Análise Discriminante de Fisher
KPCA	-	Kernel Principal Component Analysis
LFDA	-	Análise Discriminante Local de Fisher
LPP	-	Locality Preserving Projection
MANOVA	-	Análise de Variância Multivariada
NLPCA	-	Análise de Componentes Principais Não Lineares
PCA	-	Análise de Componentes Principais
PCS	-	Curvas e Superfícies Principais
SVM	-	Suport Vector Machine
UCI	-	University of California - Irwine

SUMÁRIO

1 INTRODUÇÃO

A classificação de dados amostrais em categorias é uma prática relacionada a quase todos os domínios do conhecimento humano, como educação, medicina, biologia, entre outras áreas (MINGOTI, 2005). Um médico pesquisador pode estudar a intensidade dos fatores de risco de doenças cardíacas em uma região com alta incidência e saber quais variáveis melhor preveem se um paciente é suscetível à recuperação completa (grupo 1), parcial (grupo 2) ou não (grupo 3) (HASTIE, TIBSHIRIANI e FRIEDMAN, 2009). Um biólogo pode estudar características diferentes de tipos similares (grupos) de flores e, em seguida, realizar uma análise da função discriminante para determinar o conjunto de características que permite a melhor distinção entre os tipos (FISHER, 1936).

Existem várias técnicas possíveis para a classificação de dados e a eficiência do classificador é relevante, pois minimiza a chance de classificar uma nova observação amostral incorretamente. A análise discriminante linear de Fisher (FDA) é uma técnica comumente utilizada para a classificação de dados amostrais existente em diversos *softwares* estatísticos. É uma ótima ferramenta para classificação supervisionada com muitas aplicações, devido a sua simplicidade, robustez e eficiência preditiva (HAND, 2006). Também, segundo Lim e Shih (2000) a FDA, comparada com 19 outras técnicas de classificação (NN - vizinho mais próximo, PDA - análise discriminante penalizada, POL - algoritmo polyclass, entre outras), apresentou melhor desempenho na classificação amostral, juntamente com a técnica LOG (*logistic discriminant analysis*), superior inclusive à análise discriminante quadrática. Dado a matriz *X* de dados amostrais, a FDA procura combinações lineares $Y = AX$ que melhor separem as classes de indivíduos, segundo um critério de separabilidade, minimizando a variabilidade dentro das classes e, ao fazê-lo, estará simultaneamente maximizando a dispersão entre as classes, que são previamente conhecidas no conjunto de dados observados (SHU E LU, 2014). A transformação algébrica efetuada pela análise discriminante gera um novo conjunto de pontos denominados escores discriminantes.

Aplicações da análises discriminantes são encontradas em diversas áreas: Mateus, Melo e Faria (2011) aplicaram na análise de insolvência de empresas de grande porte (sociedades anônimas), Kitani (2007) aborda o estudo das imagens de

face (classificação) como um problema de reconhecimento de padrões e Ghauri *et al.* (2014) desenvolveram um algoritmo baseado na FDA na classificação de sinais digitais. Segundo os autores, a FDA apresentou alta precisão na classificação dos dados amostrais.

Outra técnica recente para classificação é a análise discriminante local de Fisher (LFDA), desenvolvida por Sugiyama (2006) e aplicada inicialmente para redução de dimensionalidade de dados amostrais. É uma modificação da FDA e, segundo o autor, tem por objetivo preservar a estrutura local dos dados, de modo que com a transformação, os pares de pontos próximos da mesma classe continuem próximos e os pares de classes diferentes não sejam próximos. A transformação algébrica efetuada na LFDA mantém a similaridade e a dissimilaridade dos dados. Ainda segundo Sugiyama (2007), devido à propriedade de preservação da estrutura local, a LFDA incorpora melhor as características originais dos dados e é mais efetiva quando o conjunto de dados é formado por mais de dois grupos, já que na FDA as amostras são fundidas em um único grupo. Devido a essa restrição da FDA, há um menor grau de liberdade para aumentar a separabilidade, pois a LFDA não exige que o conjunto original seja transformado em um único *cluster* e como resultado, tem maior poder de separabilidade das classes.

A classificação de um novo vetor amostral, na FDA, é feita com base nas distâncias entre a observação individual e a média dos escores de cada classe (centroide). Contudo, essa metodologia pode não ser tão eficiente (FUKUNAGA, 1990; WITTEN e TIBSHIRANI, 2011). Devido às limitações da FDA e também com o objetivo de melhorar a eficiência do classificador, novas abordagens para classificação foram desenvolvidas, como por exemplo, com o uso de curvas principais. As curvas principais (CP) são uma generalização não linear das componentes principais lineares. As CPs dão um resumo da não linearidade dos dados e pode preservar melhor a adjacência dos dados (HSIEH, 2009, p. 214). A aplicação de CP na classificação de dados pode ser encontrada em diversos trabalhos: Yunsong e Huaijiang (2010) propuseram um novo classificador para dados *microarrays* usando curvas principais. Licciardi *et al.* (2012) aplicaram a NLPCA (*nonlinear principal curve analysis*) em conjunto de dados formados pelo perfil morfológico com o objetivo de investigar a precisão de classificação com estes tipos de dados. Kallas *et al.* (2012) desenvolveram um algoritmo para detecção e classificação de doenças cardíacas combinando duas técnicas – SVM (*suport vector*

machine) para classificação e KPCA (*kernel principal component analysis*) para extração das características do conjunto amostral, a qual apresentou desempenho superior ao da FDA.

Neste trabalho aplicou-se a CP na análise discriminante, com o objetivo de tornar a classificação mais eficiente que o uso dos centroides. De diversas definições para CP, é utilizado o algoritmo *k*-segmentos de Verbeek, Vlassis e Kröse (2002), denominado neste trabalho classificador *k*-segmentos. Para testar a eficiência desta técnica são utilizados 14 conjuntos de dados amostrais específicos para a análise discriminante, com 3 ou mais classes. Os conjuntos de dados amostrais utilizados neste trabalho são, na sua maioria, conjuntos do banco de dados da UCI *Learning Machines* (2014), mantido pela Universidade da Califórnia em Irvine. A UCI mantém o repositório com mais de 100 conjuntos de dados usados amplamente por estudantes, educadores e pesquisadores de todo mundo como fonte primária.

A avaliação do desempenho das técnicas aplicadas para diversas amostras é feita pela taxa de erro. Para avaliação da eficiência do método é utlizada a taxa aparente de erro (APER), que consiste em uma função para as observações amostrais que são classificadas incorretamente pela função discriminante, calculada pela reclassificação da amostra depois de obtida a regra discriminante.

A proposta deste trabalho é modificar a forma de efetuar a classificação de novas observações amostrais nas técnicas FDA e LFDA. Esta modificação é efetuada após o cálculo dos escores discriminantes, no instante em que são determinadas as distâncias do objeto a ser classificado a cada uma das classes previamente conhecidas. A FDA utiliza a distância euclidiana aos centroides de cada classe, enquanto o método proposto substitui os centroides pelas linhas poligonais desenvolvidas pelo algoritmo *k*-segmentos que também gera as curvas principais por intermédio de linhas poligonais.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Desenvolver um algoritmo para classificação amostral baseado em cada uma das técnicas FDA e LFDA, utilizando o algoritmo *k*-segmentos.

1.1.2 Objetivos específicos

a) Construir o algoritmo classificador *k*-segmentos, modificando a abordagem tradicional do método de análise discriminante, a FDA e LFDA, que efetua a classificação de novos dados amostrais através da distância ao centroide de cada classe, mostrando-se mais eficiente na classificação dos dados e apresentando melhor acurácia na classificação. Esta metodologia substitui o método dos centroides e deve apresentar melhor eficiência na classificação de dados amostrais.;
b) Implementar computacionalmente o algoritmo;
c) Avaliar o poder preditivo do algoritmo em diversos conjuntos amostrais;
d) Avaliar a eficiência discriminatória do algoritmo, comparando com as técnicas discriminantes FDA e LFDA, usando o método dos centroides.

1.2 JUSTIFICATIVA

Muitas técnicas têm sido desenvolvidas para a classificação supervisionada, que vão de métodos clássicos, como a FDA, a métodos mais recentes, tais como redes neurais, método dividir e conquistar (GUO *et al.*, 2015) e SVM (HAND, 2006). Essas ferramentas são utilizadas em várias pesquisas de diferentes áreas, como na detecção e diagnóstico de falhas em processos químicos (CHIANG, RUSSELL e

BRAATZ, 2000), na mineração de dados (*data mining*) para grandes repositórios de dados (LAST, ZHMUDYAK e SHMUELI, 2014), processamento de sinais e compressão de dados (ANARAKI e HUGHES, 2014) e bioestatística (PANG e TONG, 2012).

Outra grande ênfase para a análise discriminante no meio científico é realizar modificações na técnica FDA para melhorar a sua eficiência classificatória. Alguns trabalhos com modificações na FDA são citados a seguir: Okwonu (2014) propôs uma técnica de classificação de filtro linear para conjuntos de dados não normalmente distribuídos e que, nesta condição, apresentou desempenho robusto em relação a FDA. Hastie e Tibshirani (1996) desenvolveram uma técnica baseada na FDA para conjuntos com grandes dimensões, fazendo modificações nas matrizes de covariância S_B (entre grupos) e S_W (dentro dos grupos). Sugiyama (2006) também redefiniu as matrizes de covariância S_B e $S_{W,}$ multiplicando-as por uma matriz de afinidade. Witten e Tibshirani (2011) propuseram FDA penalizada, técnica que consiste em penalizar os vetores discriminantes, de forma que leve a uma maior interpretabilidade dos dados. Ge e Fan (2013) desenvolveram um método iterativo baseado em KFDA (*Kernel Fisher Discriminant Analysis*) para classificação de padrões. Ramli, Ismail e Wooi (2014) combinaram a técnica de classificação *k*-vizinho mais próximo com LAD (*Tree Through Stacking*) em dois tipos diferentes de dados: os macroeconômicos e risco país (estes dados foram coletados de 27 países), com o objetivo de prever o risco de crise econômica em um país.

Embora a FDA seja uma técnica de classificação com ótimo poder preditivo, em algumas situações não é tão eficiente ou não é aplicável. Lim e Shih (2000) testaram 20 algoritmos diferentes, com ênfase na precisão dos algoritmos e também no tempo de processamento dos dados e uma das conclusões obtidas é que nenhum dos algoritmos testados é uniformemente mais preciso que outro.

Considerando um conjunto amostral com *n* observações e *p* variáveis, as limitações da FDA podem ocorrer quando (FUKUNAGA, 1990; WITTEN e TIBSHIRANI, 2011):

a) A estimativa da matriz de covariância de cada classe pode ser aproximadamente singular (se *p* é quase tão grande quanto *n*);
b) O cálculo da distância euclidiana para cada centroide, bem como classificá-la no grupo que apresentar a menor distância pode penalizar

classes com menor variância e consequentemente apresentar maior probabilidade de classificação em outras classes com maior variância;

c) *p* é grande e assim a função discriminante resultante é composta de todas as *p* variáveis, o que torna difícil a interpretação;
d) Quando a amostra é formada por várias classes e com a transformação algébrica do conjunto, essas amostras são fundidas em um único conjunto para a obtenção dos escorres discriminantes, apresentando consequentemente menor eficiência na classificação;
e) A dispersão longitudinal (linear ou não linear) dos escores discriminantes pode afetar o desempenho da FDA e LFDA.

Modificações na técnica FDA podem contribuir para a melhor eficiência na classificação de dados. A utilização da menor distância dos centroides de cada classe como critério para classificação, pode não trazer bons resultados se a dispersão dos dados dentro da classe for grande.

1.3 ESTRUTURA DA TESE

A estrutura de capítulos desta tese está organizada da seguinte forma: No próximo capítulo, denominado capítulo 2, apresenta-se a revisão bibliográfica das principais técnicas utilizadas neste trabalho: a análise discriminante de Fisher (FDA), a análise discriminante local de Fisher (LFDA) e os componentes principais lineares e não lineares. No capítulo 3 descreve-se o algoritmo desenvolvido para classificação. No capítulo 4 é feita a análise dos resultados obtidos com esta nova abordagem para classificação de dados amostrais. Finalmente, no capítulo 5, conclusão e possibilidades de trabalhos futuros.

2 REVISÃO DE LITERATURA

O ser humano é capaz de reconhecer formas e padrões desde os seus primeiros anos de vida, mesmo que ainda não conheça totalmente o mecanismo que o faz reconhecer essas formas e padrões. Tal habilidade permite-lhe que seja um dos melhores classificadores existentes. Porém, muitos dos conjuntos de observações amostrais possuem diversas variáveis ou diversas observações, o que dificulta grandemente a classificação de uma nova observação amostral, pela experiência ou pela simples análise do pesquisador. Sem uma técnica formal, o procedimento de classificação pode não ter resultados satisfatórios.

O cérebro consegue analisar apenas uma quantidade limitada de variáveis ao mesmo tempo. Por exemplo, na análise de dados de um cliente pessoa física para liberação de crédito, o gerente pode ter dificuldade com a análise de diversas informações cadastrais (variáveis). Para essa análise, o gerente pode utilizar as seguintes variáveis: idade, renda, tempo de serviço na empresa, renda familiar, total de bens, número de bens, profissão, número de dependentes, etc. Consequentemente o analista deverá considerar apenas as variáveis 'mais importantes' para essa análise, com a probabilidade maior de cometer um erro na classificação. Procedimentos formais de classificação têm ajudado empresas e pesquisadores em diversas áreas de sua atuação para este fim: Médicos em diagnósticos preliminares de doenças, com o propósito de selecionar tratamentos imediatos enquanto se aguarda resultados definitivos de exames, ou no diagnóstico de falhas do processo de indústria química, com o objetivo de garantir a produtividade e a qualidade do produto (ZHAO e LI, 2012). O risco na tomada de decisão pode ser minimizado quando é utilizada uma técnica de classificação, como FDA e LFDA. A análise discriminante (tanto FDA como LFDA) é empregada para descobrir características que distinguem membros de diferentes grupos, de modo que, conhecidas as características de um novo indivíduo, seja possível prever a que grupo pertence.

Neste capítulo serão abordadas as técnicas utilizadas ou relacionadas ao assunto. Inicialmente, na seção 2.1, é abordado o conceito de análise discriminante, tema principal. Na seção 2.2 são apresentados os conceitos da LFDA. A análise de componentes principais lineares e não lineares são abordadas na seção 2.3.

2.1 ANÁLISE DISCRIMINANTE LINEAR DE FISHER

2.1.1 Introdução

Empresas, órgãos de governos e de pesquisa necessitam analisar e interpretar grandes quantidades de dados, com diversas variáveis na tomada de decisão. As informações contidas nestes conjuntos não são triviais e, para obtê-las, é necessário o uso de ferramentas adequadas para transformar os amostrais em um novo conjunto, de forma que a extração dessas informações seja relevante. A estatística multivariada possui diferentes técnicas para manipulação de dados com diversas variáveis para múltiplos fins, como a técnica de análise discriminante, que pode auxiliar na coleta de informações do conjunto de dados para classificação e discriminação (RENCHER, 2002).

A análise discriminante linear de Fisher (FDA) foi introduzida por Fisher, que aplicou a técnica no estudo do conjunto de dados Iris para três grupos (FISHER, 1936). A proposta de Fisher é transformar as observações amostrais, por meio de combinações lineares dessas variáveis em observações univariadas, de tal forma que as variáveis transformadas se apresentem o mais separado possível.

A FDA é semelhante à MANOVA (*multivariate analysis of variance*) no sentido reverso. Na MANOVA, as variáveis independentes são os grupos e a variáveis dependentes são os preditores. Na FDA, as variáveis independentes são preditoras e as variáveis dependentes são os grupos. Ela responde a pergunta: pode uma combinação linear de variáveis ser usada para prever a que grupo pertence uma nova observação experimental? Normalmente, múltiplas variáveis são incluídas na análise para ver quais as que contribuem para a discriminação entre grupos. A FDA deve ser aplicada a conjuntos amostrais em que o tamanho da amostra é maior que o número de variáveis. É usada em situações em que as classes ou grupos são conhecidos *a priori* e tem por um dos objetivos classificar uma observação, ou várias observações, nesses grupos conhecidos.

A técnica FDA é divida em duas etapas: (1) teste da importância das funções discriminantes; e (2) Classificação. O primeiro passo é idêntico ao aplicado na MANOVA. Há uma matriz de covariância total. Do mesmo modo, existem as

matrizes de covariância entre os grupos (S_B) e dentro dos grupos (S_W). Essas duas matrizes são comparadas por meio do teste F para determinar se existem ou não diferenças significativas entre grupos. Se existe tal diferença, prossegue-se com a construção das combinações lineares para classificação (se este for o objetivo).

Diversas aplicações podem ser encontradas para FDA (RANCHER 2002; JOHNSON e WICHERN, 1998: SUN, LI e SUN, 2014):

a) A comissão do processo seletivo da universidade quer classificar os candidatos mais prováveis a concluir o curso de graduação. As variáveis disponíveis são as notas no Ensino Médio, os resultados da prova do vestibular, nota do enem, etc. Com base nas variáveis do candidato, a FDA pode ser utilizada para auxiliar a universidade nesta classificação.
b) Um psiquiatra dá uma bateria de testes de diagnóstico, de modo a atribuir se um paciente é doente mental ou não. Podem-se construir as funções discriminantes para efetuar a classificação de um paciente.
c) Na classificação de crédito, um banco precisa decidir se quer ou não dar o empréstimo para um determinado cliente, isto é, o banco precisa saber se o seu cliente é ou não um cliente que pode honrar o pagamento em um tempo determinado. O banco precisa saber quais são as chances de receber o dinheiro emprestado com juros e correção. As classes são: cliente bom pagador e cliente mau pagador. As variáveis utilizadas podem ser: renda, tempo de serviço no emprego atual, tempo de conta corrente, etc.
d) Reconhecimento facial: O processo de reconhecimento facial pode ser dividido nas seguintes etapas: Detecção de faces; extração de características e representação da face; e, reconhecimento e verificação. A imagem digitalizada é representada por um grande número de pixels (matriz m x n). A FDA é aplicada na extração de características significativas. Essas regiões podem ser globais ou locais e podem ser distinguidas por textura, formas, intensidades e outras. A solução para o problema de extração de características se baseia na redução de dimensionalidade de um espaço, por intermédio de funções discriminantes.
e) Marketing: A FDA pode ser usada para determinar os fatores que distinguem os diferentes tipos de clientes para um produto.

f) Biomedicina: A principal aplicação da FDA na medicina é a avaliação do estado de gravidade de um paciente e o prognóstico da doença. Funções discriminantes são construídas para ajudar na classificação da doença de um paciente, como por exemplo, em forma leve, moderada ou grave. Na biologia, funções discriminantes podem ser úteis para a classificação de diferentes espécies de planta.

2.1.2 Análise discriminante de Fisher (FDA)

Se a solução, para determinar uma regra de classificação é reduzir a dimensionalidade do conjunto X para unidimensional apenas projetando os dados *p*-dimensionais sobre uma linha, o resultado pode apresentar uma mistura confusa de amostras de todas as classes e, portanto, com resultado não desejado. No entanto, com a rotação do conjunto é possível encontrar uma orientação para as amostras projetadas sobre a reta de modo que os dados estejam bem separados (DUDA, HART e STORK, 2001).

Dado o conjunto $X \in \mathbb{R}^{n \times p}$ descrito por p variáveis $\underline{x}^T = [x_1, x_2, ..., x_p]$ e *n* observações, dividido em *k* classes ou grupos $\pi_1, \pi_2, \ldots, \pi_k$, cada observação amostral $\underline{x}_i$ está associada a um dos elementos de $\pi \in \{\pi_1, \pi_2, \ldots, \pi_k\}$, descrevendo a que classe pertence. É necessário encontrar uma combinação linear para as componentes, tal que:

$$Y = W^T X \qquad (1)$$

De forma que sejam obtidas n amostras $y_1, \ldots, y_n$. Se *k* = 2, tem-se duas classes Y_1 e Y_2, de modo que cada $\underline{y_i}$ é a projeção da observação $\underline{x_i}$ correspondente sobre uma linha com a direção de $\underline{w_i}$. A direção de $\underline{w_i}$ é importante, no entanto também é importante que as projeções sobre a linha separem as classes com melhor eficiência, conforme mostrado na figura 1.

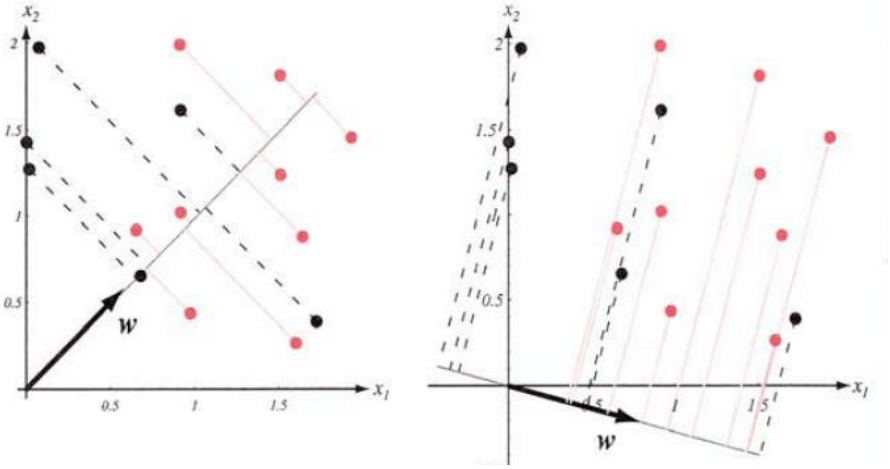


FIGURA 1 - PROJEÇÃO DA AMOSTRA SOBRE DUAS DIFERENTES LINHAS NA DIREÇÃO DE W
FONTE: BENVENISTE (2010)

O próximo passo é encontrar a direção de $\underline{w}$, com o objetivo de que o método permita a classificação com melhor eficiência. Se $\underline{\overline{x}}_i$ é o vetor médio da classe π_i com n_i observações, então:

$$\underline{\overline{x}}_i = \frac{\sum_{x \in \pi_i} x}{n_i} \tag{2}$$

E a média dos pontos projetados é dada por

$$\underline{\tilde{\overline{x}}}_i = \frac{1}{n_i}\sum_{y \in Y_i} y$$

$$= \frac{1}{n_i}\sum_{x \in \pi_i} \underline{w}^T x = \underline{w}^T \underline{\overline{x}}_i \tag{3}$$

que é a projeção de $\underline{\overline{x}}_i$. A distância entre as projeções médias é dada por

$$\left|\underline{\tilde{\overline{x}}}_1 - \underline{\tilde{\overline{x}}}_2\right| = \left|\underline{w}^T(\underline{\overline{x}}_1 - \underline{\overline{x}}_2)\right| \tag{4}$$

Esta diferença deve ser a melhor possível para obter uma boa separação dos dados. Assim,considera-se que a diferença entre as médias deve ser máxima em relação aos desvios padrões das classes. Portanto, define-se a matriz de covariância das amostras projetadas, associadas a π_i:

$$\tilde{s}_i^2 = \sum_{y \in Y_i}(\underline{y} - \underline{\tilde{\overline{x}}}_i)^2 \tag{5}$$

Isto posto, a matriz de covariância total dentro da classe das amostras projetadas é dada pela estimativa da variância dos dados agrupados,

$$\frac{\tilde{s}_1^2 + \tilde{s}_2^2}{n} \tag{6}$$

e $\tilde{s}_1^2 + \tilde{s}_1^2$ é denominada variância total dentro das classes das amostras projetadas. Uma função para quantificar significativamente a separação entre duas classes pode ser definida como:

$$J(w) = \frac{\left|\tilde{\underline{\overline{x}}}_1 - \tilde{\underline{\overline{x}}}_2\right|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \tag{7}$$

Embora a função (7) conduza à melhor separação entre as duas classes, é necessário um critério para escolher a função discriminante com melhor eficiência na classificação. Primeiro, é necessário escolher o ótimo *w* (a melhor direção) e, para obter esta função, definem-se as matrizes de covariância S_i, a matriz de covariância dentro da classe S_w e a matriz de covariância entre as classes S_B.

$$S_i = \sum_{x \in \pi_i} \left(\underline{x} - \underline{\overline{x}}_i\right)(\underline{x} - \underline{\overline{x}}_i)^T \tag{8}$$

$$S_W = S_1 + S_2 \tag{9}$$

$$S_B = \left(\underline{\overline{x}}_1 - \underline{\overline{x}}_2\right)(\underline{\overline{x}}_1 - \underline{\overline{x}}_2)^T \tag{10}$$

Substituindo em (8), tem-se:

$$S_i = \sum_{x \in \pi_i} (w^T \underline{x} - w^T \underline{\overline{x}}_i)^2 \tag{11}$$

$$= \sum_{x \in \pi_i} w^T (\underline{x} - \underline{\overline{x}}_i)(\underline{x} - \underline{\overline{x}}_i)^T w \tag{12}$$

$$= w^T S_i w \tag{13}$$

Dessa forma, a soma das variâncias pode ser escrita como:

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^T S_w w \tag{14}$$

Onde

$$S_w = \sum_{i=1}^{k} \sum_{\underline{x} \in \pi_i} (\underline{x} - \underline{\overline{x}}_i)(\underline{x} - \underline{\overline{x}}_i)^T \tag{15}$$

A separação das médias de projeção é dada por:

$$\left(\tilde{\underline{\overline{x}}}_1 - \tilde{\underline{\overline{x}}}_2\right)^2 = (w^T \underline{\overline{x}}_1 - w^T \underline{\overline{x}}_2)^2 \tag{16}$$

$$= w^T (\underline{\overline{x}}_1 - \underline{\overline{x}}_2)(\underline{\overline{x}}_1 - \underline{\overline{x}}_2)^T w$$

$$= w^T S_B w \tag{17}$$

Onde $S_B = (\underline{\overline{x}}_1 - \underline{\overline{x}}_2)(\underline{\overline{x}}_1 - \underline{\overline{x}}_2)^T$

A FDA é uma técnica para a classificação supervisionada com o objetivo de encontrar uma orientação para que as amostras projetadas sejam bem separadas, o que é equivalente encontrar uma projeção que minimiza a distância entre os centros de cada classe. Devido ao fato de que a dispersão em torno dos centros venha a desempenhar um papel importante, não é suficiente considerar apenas os centros

das classes, mas também a matriz de covariância de cada classe. Assim, é importante maximizar a distância entre as classes e minimizar a distância dentro das classes simultaneamente, ou seja, o critério de Fisher descreve a separabilidade em termos de S_B e S_w, que pode ser escrito como:

$$J(w) = \frac{w^T S_B w}{w^T S_w w} \tag{18}$$

Onde $\tilde{s}_1^2 + \tilde{s}_1^2 = w^T S_w w$ e $\left(\tilde{\underline{\bar{x}}}_1 - \tilde{\underline{\bar{x}}}_2\right)^2 = w^T S_B w$.

Para encontrar o máximo para $J(w)$, determina-se a derivada de $J(w)$ e iguala-se a zero:

$$\frac{dJ(w)}{dw} = \frac{d}{dw}\left[\frac{w^T S_B w}{w^T S_w w}\right] = 0 \tag{19}$$

$$[w^T S_w w]\ \frac{d}{dw}[w^T S_B w] - w^T S_B w \frac{d}{dw}[w^T S_w w] = 0 \tag{20}$$

$$[w^T S_w w]\ 2S_B w - w^T S_B w\ 2S_w w = 0 \tag{21}$$

Dividindo por $w^T S_w w$

$$\frac{w^T S_B w}{w^T S_w w} S_B w - S_w w \frac{w^T S_B w}{w^T S_w w} = 0 \tag{22}$$

$$S_B w - S_w w\, J(w) = 0 \tag{23}$$

$$S_w^{-1} S_B w - J(w) w = 0 \tag{24}$$

$$S_B w = J(w) S_w w \tag{25}$$

A solução da equação (25) pode ser obtida com o problema dos autovetores generalizados, onde w é a matriz dos autovetores de $S_w^{-1} S_B$ e $J(w)$ a matriz diagonal com autovalores correspondentes (JOHNSON e WICHERN, 1998). Para duas classes a solução para W que otimiza *J(.)* é:

$$W = S_w^{-1}(\overline{x}_i - \overline{x}) \tag{26}$$

Quando as probabilidades *a posteriori* $p(x/\pi_i)$ normalmente distribuídas com matrizes de covariância $\sum$ iguais, a função discriminante é:

$$Y = S_W^{-1}(\overline{x}_i - \overline{x})X \tag{27}$$

Para o problema com *k* classes, a generalização da FDA envolve *k-1* funções discriminantes. A projeção é de um espaço de dimensão *p* para um espaço de dimensão *k*, de modo que $p \geq k$. A generalização da matriz de covariância dentro das classes é dada por:

$$S_w = \sum_{i=1}^{k} S_i \tag{28}$$

Onde

$$S_i = \sum_{x \in \pi_i} \left(\underline{x} - \underline{\overline{x}}_i\right)\left(\underline{x} - \underline{\overline{x}}_i\right)^T \quad (29)$$

e

$$\underline{\overline{x}}_i = \sum_{x \in \pi_i} \frac{x}{n_i} \quad (30)$$

A generalização de S_B é dada da seguinte forma: Considere que o vetor da média total é definido por $\underline{\overline{x}} = \sum_x \frac{x}{n} = \sum_{i=1}^{k} \frac{n_i x_i}{n}$ e a matriz de covariância total

$$S_T = \sum_x (\underline{x} - \underline{\overline{x}})(\underline{x} - \underline{\overline{x}})^T \quad (31)$$

Segue que

$$s_T = \sum_{i=1}^{k} \sum_{x \in \pi_i} (\underline{x} - \underline{\overline{x}}_i + \underline{\overline{x}}_i - \underline{\overline{x}})\,(\underline{x} - \underline{\overline{x}}_i + \underline{\overline{x}}_i - \underline{\overline{x}})^T \quad (32)$$

$$= \sum_{i=1}^{k} \sum_{x \in \pi_i} \left(\underline{x} - \underline{\overline{x}}_i\right)\left(\underline{x} - \underline{\overline{x}}_i\right)^T + \sum_{i=1}^{k} \sum_{x \in \pi_i} \left(\underline{\overline{x}}_i - \underline{\overline{x}}\right)\left(\underline{\overline{x}}_i - \underline{\overline{x}}\right)^T \quad (33)$$

$$= S_w + \sum_{i=1}^{k} \left(\underline{\overline{x}}_i - \underline{\overline{x}}\right)\left(\underline{\overline{x}}_i - \underline{\overline{x}}\right)^T \quad (34)$$

Finalizando

$$S_T = S_W + S_B \quad (35)$$

Onde $S_B = \sum_{i=1}^{k} \left(\underline{\overline{x}}_i - \underline{\overline{x}}\right)\left(\underline{\overline{x}}_i - \underline{\overline{x}}\right)^T$.

As amostras $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ projetadas por um conjunto correspondente de amostras $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$, podem ser descritas pelos vetores da média e da matriz de covariância.

$$\underline{\tilde{\overline{x}}}_i = \sum_{y \in Y_i} \frac{y}{n_i} \quad (36)$$

$$\underline{\tilde{\overline{x}}} = \sum_{i=1}^{k} \frac{n_i \underline{\tilde{\overline{x}}}_i}{n} \quad (37)$$

$$\tilde{S}_W = \sum_{i=1}^{k} \sum_{y \in Y_i} \left(\underline{y} - \underline{\tilde{\overline{x}}}_i\right)\left(\underline{y} - \underline{\tilde{\overline{x}}}_i\right)^T \quad (38)$$

e

$$\tilde{S}_B = \sum_{i=1}^{k} n_i \left(\underline{\tilde{\overline{x}}}_i - \underline{\tilde{\overline{x}}}\right)\left(\underline{\tilde{\overline{x}}}_i - \underline{\tilde{\overline{x}}}\right)^T, \quad (39)$$

Consequentemente, tem-se:

$$\tilde{S}_W = W^T S_W W \quad (40)$$

e

$$\tilde{S}_B = W^T S_B W \quad (41)$$

As equações (40) e (41) mostram que o problema de extração das características dos dados originais pode ser transferido para o problema de encontrar uma projeção para W que otimize o critério de Fisher, dado pela equação

(42). Um dado importante é que o critério de Fisher é bastante atraente, pois é mais fácil distinguir um grupo de outro se a soma dos quadrados entre classes para *Y*= $W^T X$ é grande com relação à soma dos quadrados dentro das classes (FERREIRA, 2008). A razão entre a soma dos quadrados dentro e entre as *k* classes é dada por:

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \quad (42)$$

Se W é a matriz que maximiza a expressão (42), então, *Y* = $W^T X$ é a matriz das funções discriminantes de Fisher e W é a matriz dos autovetores de $S_W^{-1} S_B$. Segue o teorema e a prova dessa conclusão (JOHNSON e WICHERN, 1998, p. 654-655).

Teorema: Sejam S_w e S_B duas matrizes simétricas de ordem p, então o vetor que maximiza a razão $\frac{\underline{w}^T S_B \underline{w}}{\underline{w}^T S_w \underline{w}}$ é dado pelo autovetor de $S_w^{-1} S_B$ correspondente ao maior autovalor.

Demonstração: O vetor $\underline{a}$ pode ser escolhido arbitrariamente, desde que não afete a razão $\frac{\underline{a}^T S_B \underline{a}}{\underline{a}^T S_W \underline{a}}$. Assim, pode-se reformular o problema da seguinte forma: é necessário encontrar o vetor $\underline{a}$ que maximiza $\underline{a}^T S_B \underline{a}$ sujeito a seguinte restrição: $\underline{a}^T S_W \underline{a}$= 1.

Seja ${S_W}^{1/2}$ a matriz raiz quadrada de S_W. Seja $\underline{z} = {S_W}^{1/2} \underline{a}$ tal que $\underline{a} = {S_W}^{-1/2} \underline{z}$. Então

$$\underline{a}^T S_B \underline{a} = ({S_W}^{-1/2} \underline{z})^T S_B \, ({S_W}^{-1/2} \underline{z}) = \underline{z}^T {S_W}^{-1/2} S_B {S_W}^{-1/2} \underline{z} \quad (43)$$

e

$$\underline{a}^T S_W \underline{a} = ({S_W}^{-1/2} \underline{z})^T S_W \, ({S_W}^{-1/2} \underline{z}) = \underline{z}^T \underline{z}. \quad (44)$$

Logo, o máximo de $\underline{a}^T S_B \, \underline{a}$ sujeito a restrição $\underline{a}^T S_W \underline{a}$= 1 é o mesmo que o máximo de $({S_W}^{-1/2} \underline{z})^T S_B \, ({S_W}^{-1/2} \underline{z})$ sujeito à restrição $\underline{z}^T \underline{z}$ = 1.

Considere a matriz ${S_W}^{-1/2} S_B {S_W}^{-1/2}$. Essa matriz é simétrica e, portanto, pelo teorema da decomposição espectral de uma matriz simétrica, ela pode ser escrita na forma $P \Lambda P^T$, onde Λ é uma matriz diagonal formada com os autovalores de ${S_W}^{-1/2} S_B {S_W}^{-1/2}$ em ordem decrescente ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$) e *P* é uma matriz ortogonal cujas colunas são os autovetores padronizados de ${S_W}^{-1/2} S_B {S_W}^{-1/2}$. Agora, seja $\underline{w} = P^T \underline{z}$, consequentemente $\underline{z} = P \underline{w}$, pois $PP^T = I$.

Então

$$\underline{z}^T {S_W}^{-1/2} S_B {S_W}^{-1/2} \underline{z} = \underline{z}^T P \Lambda P^T \underline{z} = \underline{w}^T \Lambda \underline{w} \quad (45)$$

e

$$\underline{z}^T \underline{z} = \underline{z}^T P P^T \underline{z} = \underline{w}^T \underline{w}. \quad (46)$$

O problema pode ser reformulado novamente: é necessário encontrar um vetor $\underline{w}$ que maximize $\underline{w}^T \Lambda \underline{w} = \sum_{i=1}^{p} \lambda_i w_i^2$ sujeito à restrição $\underline{w}^T \underline{w} = 1$.

Considerando que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$, tem-se que $max_w \sum_{i=1}^{p} \lambda_i w_i^2 \leq \lambda_1 max_w \sum_{i=1}^{p} w_i^2$, e a igualdade $\underline{w}^T \underline{w} = 1$ vale para todo vetor $\underline{w}$. Além disso, o valor máximo para λ_1 é atingido quando $\underline{w}^T = (1, 0, \ldots, 0)$. Assim, o vetor $\underline{w}$ que maximiza $\underline{w}^T \Lambda \underline{w}$ sujeito a restrição $\underline{w}^T \underline{w} = 1$ é $\underline{w}^T = (1, 0, \ldots, 0)$. Logo, isso significa que $\underline{z} = P\underline{w} = \underline{v}_1$ e $\underline{a} = {S_W}^{-1/2} \underline{z} = \underline{a} = {S_W}^{-1/2} \underline{v}_1$, onde $\underline{v}_1$ é o autovetor correspondente ao autovalor λ_1.

Considerando duas matrizes $A_{nx\,p}$ e $C_{p\,x\,n}$, tal que os autovetores não nulos de AC e de CA são os mesmos e tem a mesma multiplicidade. Fazendo $A = {S_W}^{-1/2} S_B$ e $C = {S_W}^{-1/2}$, significa que os autovalores de $AC = {S_W}^{-1/2} S_B {S_W}^{-1/2}$ são os mesmos autovalores de $CA = {S_W}^{-1} S_B$. Assim, λ_1 é também o maior autovalor de ${S_W}^{-1} S_B$ e $\underline{v}_1$ é o autovetor correspondente ao maior autovalor λ_1 de ${S_W}^{-1/2} S_B {S_W}^{-1/2}$, logo:

$${S_W}^{-1} S_B ({S_W}^{-\frac{1}{2}} \underline{v}_1) = {S_W}^{-\frac{1}{2}} \left({S_W}^{-\frac{1}{2}} S_B {S_W}^{-\frac{1}{2}} \underline{v}_1 \right) = {S_W}^{-\frac{1}{2}} \lambda_1 \underline{v}_1 = \lambda_1 \left({S_W}^{-\frac{1}{2}} \underline{v}_1 \right). \quad (47)$$

Assim, tem-se que $\underline{a} = {S_W}^{-1/2} \underline{v}_1$ é o autovetor de ${S_W}^{-1} S_B$ correspondente ao maior autovalor λ_1.

2.1.3 Taxa de Erro Aparente (APER)

Dada uma regra de classificação, o desempenho da mesma deve ser avaliado pela sua capacidade de bem classificar ou de mal classificar. A taxa de erro ou taxa de má-classificação é, segundo Webb (2002), a taxa mais comumente utilizada para avaliar o desempenho de um classificador. Essa taxa representa a proporção ou porcentagem de padrões classificados incorretamente. Conjuntamente à taxa de erro é construída a matriz de confusão ou matriz de má-classificação. Cada elemento dessa matriz representa o número de padrões da classe *j* que foram

classificados como classe *i*. A APER é uma função para as observações amostrais que são classificadas incorretamente pela função discriminante. A fórmula da APER é:

$$\text{APER} = \frac{\sum_{i=1}^{K} n_{iM}}{\sum_{i=1}^{k} N_i} = \frac{n_{1M}+n_{2M}+\cdots+n_{kM}}{N_1+N_2+\cdots+N_k} \quad (48)$$

Onde n_{iM} denota o número de classificações incorretas na *i*-ésima classe e N_i denota o número de observações da *i*-ésima classe.

A APER é intuitivamente atraente e fácil de calcular. Infelizmente, tende a subestimar a taxa real de erro - AER (TIMM, 2002). Outra abordagem não paramétrica que pode funcionar melhor é a de Lachenbruch e Mickey (1968). Esta é denominada *leave-one-out* ou validação cruzada. A técnica consiste em:

1. Inicie com a primeira classe π_1. Retire uma das observações de π_1 e determine a função discriminante baseada nas $N_1 - 1,\ N_{2,} \ldots, N_k$ observações;
2. Classifique a observação retirada na etapa 1;
3. Repita os passos 1 e 2 para todas as observações de π_1. Seja n_{1M}^{H} o número de classificações incorretas no grupo π_1;
4. Repita também os passos de 1 a 3 para as observações da classe 2. Seja n_{2M}^{H} o número de classificações incorretas no grupo π_2. Repita o mesmo procedimento para as classes restantes;

A taxa de erro é dada por:

$$E(AER) = \frac{\sum_{i=1}^{k} n_{iM}^{H}}{\sum_{i=1}^{k} N_i} \quad (49)$$

Conjuntamente com a APER utiliza-se a matriz de confusão, apresentada na tabela 1, para 2 classes.

TABELA 1 - MATRIZ DE CONFUSÃO

Classe Real	Classe atribuída		Tamanho da classe
	π_1	π_2	
π_1	n_{1c}	n_{1M}	N_1
π_2	n_{2M}	n_{2c}	N_2

FONTE: JOHNSON e WICHERN (1998)

Onde:

n_{1c} é o número de indivíduos de π_1 classificado corretamente em π_1.

n_{2c} é o número de indivíduos de π_2 classificado corretamente em π_2.
n_{1M} é o número de indivíduos de π_1 classificado incorretamente em π_2.
n_{2M} é o número de indivíduos de π_2 classificado incorretamente em π_1.

A matriz de confusão e a APER são utilizadas para justificar o quanto a regra de classificação foi eficiente. Ainda, a APER é uma estimativa da probabilidade de que o processo de classificação irá considerar incorretamente uma nova observação.

2.2 LFDA

A LFDA (análise discriminante local de Fisher) combina duas técnicas, a FDA e a LPP (*locality preserving projection*). Para SUGIYAMA (2007), a transformação algébrica pela FDA pode não ser adequada quando o conjunto de dados é composto por várias classes (três ou mais). O comportamento indesejado é causado pela globalidade ao avaliar a dispersão dentro e entre as classes. A preservação da estrutura local dos dados é apontada como solução para esse problema, isto é, se a amostra $\underline{x}_i$ é próxima de $\underline{x}_j$ e vice-versa, com a transformação algébrica, a proximidade do par de dados deve ser mantida no novo espaço algébrico. A LPP transforma a matriz de dados de forma que os pares de dados próximos no espaço original são mantidos próximos no espaço gerado. Mas a LPP também faz a transformação globalizada (não supervisionada). Como solução dessas limitações, Sugiyama combinou a FDA com LPP.

Seja *X* a matriz de todas as amostras, de ordem *n x p,* usando a matriz de transformação *T* (matriz dos pesos discriminantes para LFDA)*,* a matriz *Z* é dada por $Z = T^T X$.

2.2.1 LPP - *locality preserving projection*

Seja $A_{n\ x\ n}$ uma matriz de afinidade, isto é, o elemento a_{ij} como sendo a afinidade entre as observações amostrais $\underline{x}_i$ e $\underline{x}_j$, um modo simples de representar os

elementos da matriz A (não é a única forma) é definir a_{ij} = 1, se $\underline{x}_j$ é *k*-vizinho mais próximo de $\underline{x}_j$ ou vice-versa; caso contrário, a_{ij} = 0.

Dado A, a matriz de transformação T_{LPP} é definida da seguinte maneira:

$$T_{LPP} = argmin_{T \in \mathbb{R}^{n \times p}} \frac{1}{2} \sum_{i,j}^{n} A_{ij} \left\| T^T \underline{x}_i - T^T \underline{x}_j \right\|^2 \quad (50)$$

Sujeita a $T^T X D X^T T = I$, onde *I* é uma matriz identidade e *D* é uma matriz diagonal com o i-ésimo elemento dado por:

$$d_{i,i} = \sum_{j=1}^{n} a_{ij}. \quad (51)$$

Essa transformação determina que um par de pontos próximos no conjunto original é mantido próximo após a transformação. A restrição $T^T X D X^T T = I$ dessa transformação é para evitar a solução trivial *T = 0*.

A matriz de transformação T_{LPP} é dada por:

$$T_{LPP} = (\underline{e}_{p-m+1} | \underline{e}_{p-m+2} | \dots | \underline{e}_p), \quad (52)$$

Onde $\underline{e}_j$ são os autovetores associados aos seus autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ para o seguinte problema de autovalores:

$$X L X^T \underline{e} = \lambda X D X^T \underline{e} \quad (53)$$

Onde $L = D - A$.

2.2.2 Análise Discriminante Local de Fisher – LFDA

Conforme pode ser visto na figura 2 adiante, as três técnicas (FDA, LFDA e LPP) podem apresentar resultados semelhantes, com a projeção sobre as linhas semelhantes nos três métodos.

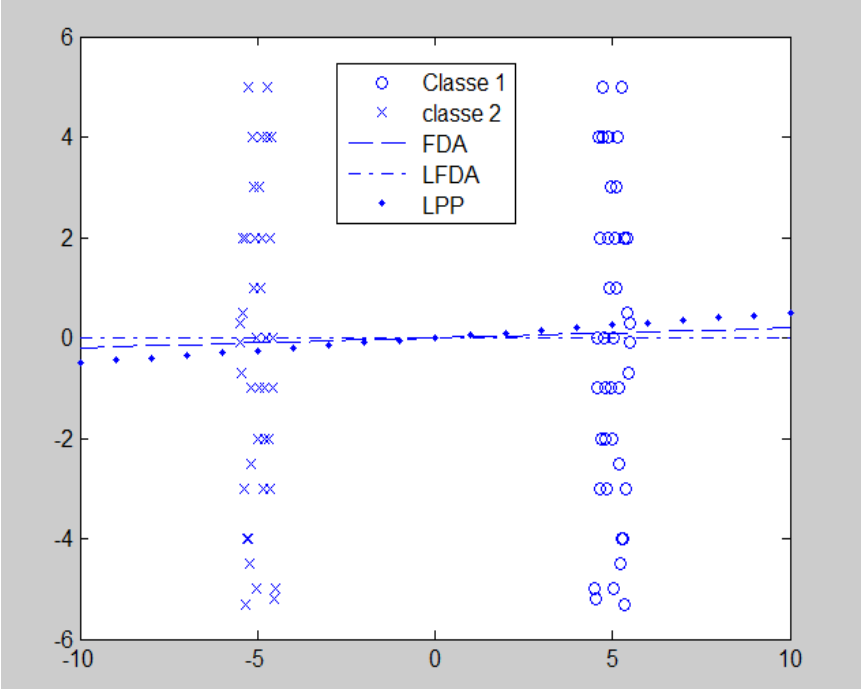
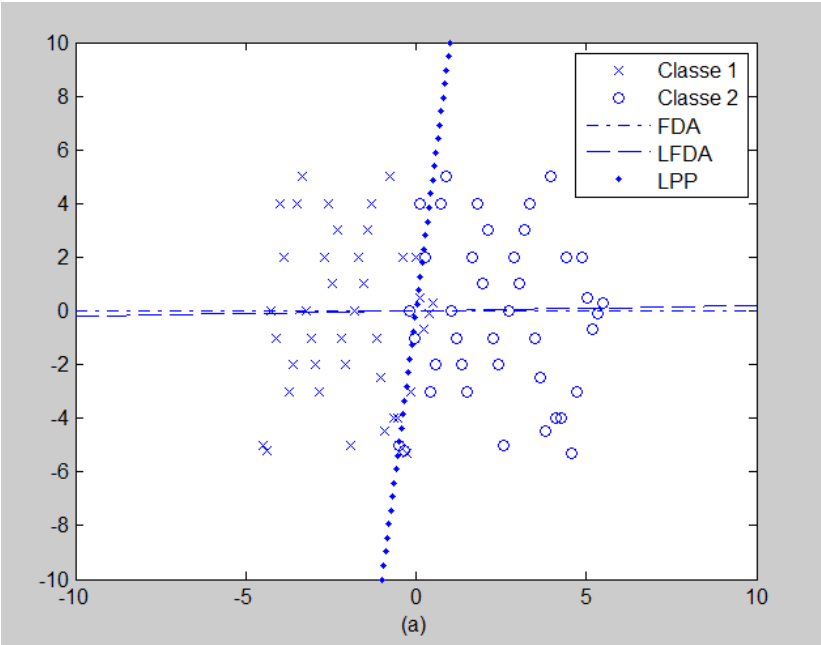


FIGURA 2 - EXEMPLO DE APLICAÇÃO DA FDA, LFDA E LPP
FONTE: SUGIYAMA (2007)

Porém, segundo Sugiyama (2007) a FDA pode executar mal a separação das classes devido à globalização dos dados, quando se avalia a variância dentro da classe e entre classes, conforme figura 3b. Por outro lado, a LPP pode também não ser eficiente na avaliação da dispersão dos dados, como na figura 3a.



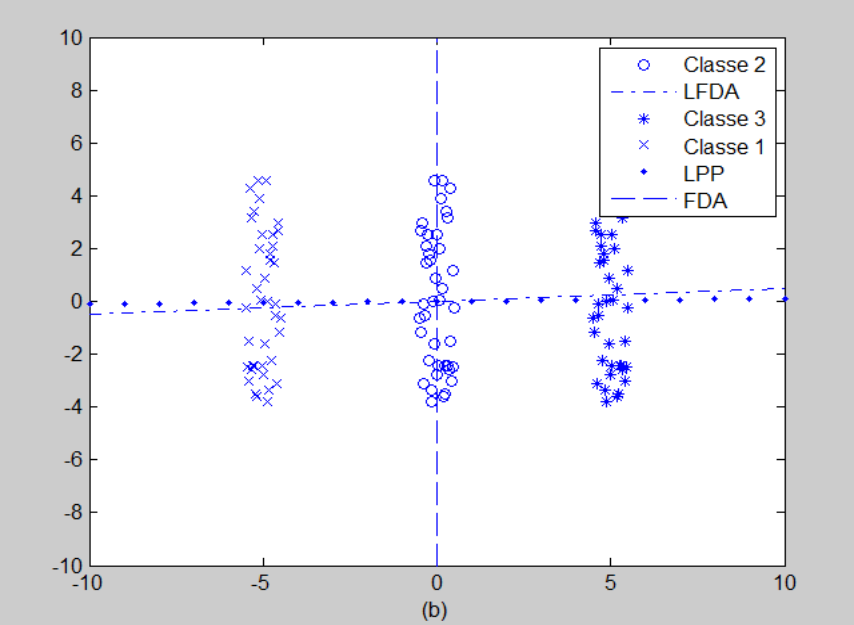


FIGURA 3 - EXEMPLO DE APLICAÇÃO DA FDA, LFDA E LPP
FONTE: SUGIYAMA (2007)

No primeiro gráfico (a) a LPP não apresentou resultado adequado e no segundo (b), a FDA.

LFDA é definida da seguinte forma: Seja a matriz de covariância local dentro da classe S_{lW} e a matriz de covariância local entre as classes S_{lB}.

$$S_{lW} = \frac{1}{2}\sum_{ij=1}^{n} A_{ij}^{w}(\underline{x}_i - \underline{x}_j)(\underline{x}_i - \underline{x}_j)^T \tag{54}$$

e $$S_{lB} = \frac{1}{2}\sum_{ij=1}^{n} A_{ij}^{B}(\underline{x}_i - \underline{x}_j)(\underline{x}_i - \underline{x}_j)^T \tag{55}$$

Onde A_{ij}^{w} e A_{ij}^{B} são matrizes de afinidades, tais que:

$$A_{ij}^{w} = \begin{cases} \frac{1}{n_c}, & se \quad y_i = y_j = c \\ 0, & se \quad y_i \neq y_j \end{cases} \tag{56}$$

e

$$A_{ij}^{B} = \begin{cases} \frac{1}{n} - \frac{1}{n_c}, & se \quad y_i = y_j = c \\ 0, & se \quad y_i \neq y_j \end{cases} \tag{57}$$

Demonstração: Da equação (15) tem-se que:

$$S_{lW} = \sum_{i=1}^{l} \sum_{j:y_j=i} (\underline{x}_j - \frac{1}{n_i} \sum_{p:y_p=i} \underline{x}_p)(\underline{x}_j - \frac{1}{n_i} \sum_{q:y_q=i} \underline{x}_q)^T \quad (58)$$

$$= \sum_{i=1}^{n} \underline{x}_i \underline{x}_i^T - \sum_{i=1}^{l} \frac{1}{n_i} \sum_{p,q:y_p=y_q=i} \underline{x}_p \underline{x}_q^T \quad (59)$$

$$= \sum_{i=1}^{n} (\sum_{j=1}^{n} A_{ij}^w) \underline{x}_i \underline{x}_i^T - \sum_{i,j=1}^{n} A_{ij}^w \underline{x}_i \underline{x}_j^T \quad (60)$$

$$= \frac{1}{2} \sum_{ij=1}^{n} A_{ij}^w (\underline{x}_i \underline{x}_i^T + \underline{x}_j \underline{x}_j^T - \underline{x}_i \underline{x}_j^T - \underline{x}_j \underline{x}_i^T) \quad (61)$$

$$S_{lW} = \frac{1}{2} \sum_{ij=1}^{n} A_{ij}^w (\underline{x}_i - \underline{x}_j)(\underline{x}_i - \underline{x}_j)^T \quad (62)$$

A dispersão total é:

$$S_{lT} = S_{lW} + S_{lB} = \sum_{i=1}^{n} (\underline{x}_i - \overline{\underline{x}})(\underline{x}_i - \overline{\underline{x}})^T \quad (63)$$

Segue que

$$S_{lB} = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i^T - \frac{1}{n} \sum_{ij=1}^{n} \underline{x}_i \underline{x}_j^T - S_{lw} \quad (64)$$

$$= \sum_{i=1}^{n} (\sum_{j=1}^{n} \frac{1}{n}) \underline{x}_i \underline{x}_i^T - \sum_{ij=1}^{n} \frac{1}{n} \underline{x}_i \underline{x}_j^T - S_{lw} \quad (65)$$

$$= \frac{1}{2} \sum_{ij=1}^{n} (\frac{1}{n} - A_{ij}^w) (\underline{x}_i \underline{x}_i^T + \underline{x}_j \underline{x}_j^T - \underline{x}_i \underline{x}_j^T - \underline{x}_j \underline{x}_i^T) \quad (66)$$

$$S_{lB} = \frac{1}{2} \sum_{ij=1}^{n} A_{ij}^B (\underline{x}_i - \underline{x}_j)(\underline{x}_i - \underline{x}_j)^T \quad (67)$$

Usando S_{lW} *e* S_{lB}, porém modificando as matrizes de afinidades:

$$A_{ij}^w = \begin{cases} \frac{A_{ij}}{n_c}, & se \quad y_i = y_j = c \\ 0, & se \quad y_i \neq y_j \end{cases} \quad (68)$$

e

$$A_{ij}^B = \begin{cases} A_{ij}(\frac{1}{n} - \frac{1}{n_c}), & se \quad y_i = y_j = c \\ 0, & se \quad y_i \neq y_j \end{cases} \quad (69)$$

A matriz de transformação T_{LFDA} é definida como:

$$T_{LFDA} = \arg\max tr\,(\frac{w^T S_{lB} w}{w^T S_{lw} w}) \quad (70)$$

Determina-se T para que os pares de dados próximos da mesma classe sejam pares próximos no novo espaço e os dados diferentes mantenham-se separados.

2.3 CURVAS PRINCIPAIS

A análise de componentes principais (PCA - *Principal Componentes Analysis*) é uma técnica da análise multivariada que encontra aplicações nas mais diversas áreas, como (HASTIE, TIBSHIRANI, FRIEDMAN (2009); SHLENS (2014); MINGOTI (2005)):

a) Médica;

b) Meteorológica;

c) Estatística;

d) Inteligência artificial;

e) Processamento de sinais;

f) Neurociência.

O objetivo da PCA é reduzir a dimensão do conjunto original, ou distribuí-lo, de modo que a nova base descreva melhor os padrões de variabilidade existentes nos dados. No início da década de 90, uma generalização da PCA utilizando redes neurais foi desenvolvida por Kramer (1991), na área de engenharia química, a qual denominou como análise componente principal não-linear (NLPCA). Outra solução a esse problema, vinda da comunidade estatística, foi proposta por Hastie (1984) e por Hastie e Stuetzle (1989), que nomearam como método das curvas e superfícies principais (PCS). A seguir é feita uma breve revisão sobre componentes principais lineares e não lineares.

2.3.1 Componentes principais lineares

A análise de componentes principais (PCA) é um método para transformar os dados originais de modo a eliminar parte da informação redundante em cada dimensão. É um método de descorrelação de dados. A PCA se baseia em uma simples transformação linear em relação ao conceito dos mínimos quadrados tal que a função custo da PCA minimiza a soma das distâncias de projeção ortogonais dos pontos sobre uma linha. O objetivo da PCA é encontrar uma transformação mais representativa e de preferência mais compacta das observações. Desse modo, os

dados podem ser descritos de forma mais concisa, tal que as primeiras poucas dimensões expliquem o maior número possível de informações disponíveis, bem como o tipo de relacionamento existente entre as variáveis.

O método de PCA transforma um vetor aleatório $\underline{x} \in \mathbb{R}^p$ em outro vetor $\underline{y} \in \mathbb{R}^k$, projetando $\underline{x}$ nas *k* direções ortogonais de maior variância. Geometricamente, os autovetores e seus autovalores da matriz de covariância Σ refletem as direções e a variação das direções dos dados. Algebricamente, $\underline{y}$ é uma combinação linear de *p* variáveis originais e, geometricamente, as combinações lineares representam a seleção de um novo sistema de coordenadas, obtido por rotação do sistema original.

Dado que as variáveis são correlacionadas, a PCA faz uma simples transformação linear, que tem a seguinte forma:

$$\underline{y} = A\underline{x} + \underline{b} \tag{71}$$

Onde $A \in \mathbb{R}^{kxp}$ e $\underline{b}$, $\underline{y} \in \mathbb{R}^k$. A matriz A é uma matriz tal que as colunas são vetores ortonormais, considerando que existe um tipo de relação entre as variáveis. A equação pode ser reescrita na forma:

$$\underline{x} = P\underline{y} + \underline{\varepsilon} \tag{72}$$

ou

$$\underline{x} = f(\underline{y}) + \underline{\varepsilon} \tag{73}$$

Onde $P \in \mathbb{R}^{Pxk}$ é uma matriz que descreve o interrelacionamento entre as variáveis de $\underline{x}$ e a variável latente $\underline{y}$ e $\underline{\varepsilon}$ representa o ruído (ou erro), tal que $E(\underline{\varepsilon}) = \underline{0}$, $E(\underline{\varepsilon}^T \underline{\varepsilon}) = \underline{\delta} I$, onde $\underline{\delta}$ é a variância do ruído e I é a matriz identidade.

O vetor de média μ_y e a matriz de covariância Σ_y são:

$$\mu_y = E\{\underline{y}\} = E\{A\underline{x} + \underline{b}\} = A\underline{\mu}_x + \underline{b} \tag{74}$$

$$\sum_y = E\{(\underline{y} - \underline{\mu}_x)(\underline{y} - \underline{\mu}_x)^T\} = A\, E\{(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^T\} A^T \tag{75}$$

$$\sum_y = A \sum_x A^T \tag{76}$$

$$\sum_x = E\{(\underline{x} - \underline{\mu}_x)(\underline{x} - \underline{\mu}_x)^T\} = E\{\underline{x}\underline{x}^T\} - \underline{\mu}_x \underline{\mu}_x{}^T \tag{77}$$

Observe que o vetor $\underline{b}$ não afeta o valor da matriz de covariância de Y, mas sim a média de $\underline{y}$. A matriz de covariância $\sum_x$ é uma matriz positiva semidefinida, isto é, os autovalores são maiores ou iguais a zero.

Se $|\sum_x| \neq 0$, $\sum_x$ pode ser decomposto no seguinte produto matricial:

$$\sum_x = \Gamma_x \Lambda_x \Gamma_x^T \tag{78}$$

Para

$$\Gamma x = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1k} \\ \vdots & \ldots & \vdots \\ \gamma_{k1} & \cdots & \gamma_{kk} \end{bmatrix} = [\underline{\gamma}_1, \ldots, \underline{\gamma}_k], \qquad \Lambda x = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix}$$

Em que $\lambda_1 \geq, \ldots, \geq \lambda_k$ são os autovalores de $\sum_x$ e $\underline{\gamma}_1, \ldots, \underline{\gamma}_k$ os autovetores correspondentes e $\underline{\gamma}_i^T \underline{\gamma}_j = 0$ para i ≠ j. A transformação linear passa a ter a seguinte forma:

$$\underline{y} = A\underline{x} = \Gamma_x^T \underline{x} \tag{79}$$

Consequentemente tem-se:

$$\sum_y = A\sum_x A^T = \Gamma_x^T \sum_x \Gamma_x = \Gamma_x^T \Gamma_x \Lambda_x \Gamma_x^T \Gamma_x = \Lambda_x \tag{80}$$

Esta transformação projeta $\underline{x}$ em direções ortogonais de modo que o vetor transformado $\underline{y}$ tenha uma matriz de covariância diagonal e $\sigma_{\underline{y}_i}^2 = \lambda_i$.

2.3.2 Componentes principais não lineares

Considere um conjunto de dados consistindo de n observações com duas variáveis, x e y. Estes n pontos podem ser representados em um sistema cartesiano, de várias formas, conforme mostra a figura 4 (página 42). A técnica de modelagem utilizada para esse conjunto depende do objetivo da análise desses dados. Se a média for utilizada para resumir esse conjunto, não se tem a informação sobre o comportamento das duas variáveis. Se o objetivo é de verificar a existência de uma relação funcional entre uma variável dependente com uma (ou mais) variável independente, pode-se utilizar a regressão linear, conforme figura 4a. Este procedimento equivale a encontrar uma linha que minimiza as somas dos desvios quadrados, com projeção vertical de cada vetor. Mas, se é necessário estudar o comportamento das duas variáveis, a regressão pode não ser uma boa técnica. Uma alternativa é resumir os dados por uma linha reta que trata as duas variáveis simetricamente. A linha definida pela primeira componente principal faz isso, projetando ortogonalmente os pontos sobre a linha, de modo que os desvios sejam mínimos, como destacado na figura 4b, que mostra a componente principal linear dos pontos projetados ortogonalmente sobre a linha.

Se a estrutura dos dados é não linear, é necessário o uso de técnicas não lineares para sumarizar o conjunto[1]. Vários métodos para obter as curvas suaves (*kernel smoothers, nearest-neighbor smoothers, spline smoothers*) podem ser aplicados a este problema. Diagramas de dispersão podem ser suavizados pela montagem de uma linha para os pontos do conjunto. Em geral, esses métodos produzem uma curva que minimiza a projeção vertical sobre a curva, conforme figura 4c. De modo similar aos métodos citados para curvas suaves, a CP é uma generalização da técnica de PCA. Em vez de resumir os dados em uma linha reta, é utilizada uma curva suave com a projeção ortogonal dos dados, conforme figura 4d. Na figura 4a, a linha de regressão linear minimiza a soma dos quadrados dos desvios verticais entre cada evento $\underline{x}$. Na 4b, a PCA é uma regressão linear simétrica (as variáveis são independentes) e é a direção que minimiza as distâncias ortogonais dos pontos à reta. Na 4c, a regressão não linear emprega funções não lineares que minimizam a distância vertical entre as variáveis. Na 4d, a CP minimiza a distância ortogonal entre os dados e a curva.

A seguir são apresentados alguns dos conceitos mais importantes para CP desenvolvida por Hastie e Stuetzle (1989) como a definição de CP e índice de projeção.

Uma curva em um espaço *p*-dimensional é uma função continua $f: I \rightarrow \mathbb{R}^p$, onde I = [a,b] com a, b $\in \mathbb{R}$.

A curva f pode ser considerada um vetor de p funções de uma única variável independente t, $f(t)$ = $(f_1(t), \ldots, f_p(t))$, onde $f_1(t), \ldots, f_p(t)$ são denominadas funções coordenadas. A curva $f(t)$ = ($f_1(t), \ldots, f_p(t)$) é uma curva em $\mathbb{R}^p$ parametrizada por $t \in \mathbb{R}$ e a distância entre $\underline{x}$ e $f(t)$, denotada por $t_f(\underline{x})$, é a menor distância de $\underline{x}$ à curva. O ponto de projeção de $\underline{x}$ sobre f é $f(t_f(\underline{x}))$. A curva $f(t)$ é uma curva principal para *X, se* $f(t)$ *é* autoconsistente*, isto é,* $f(t) = E[X \setminus t_f(\underline{x}) = t]$. O índice de projeção $t_f : \mathbb{R}^p \rightarrow \mathbb{R}$ é definido por:

[1] Algoritmo para testar a não linearidade do conjunto para Componentes principais, pode ser obtido em Krüger, Zhang e Xie (2008).

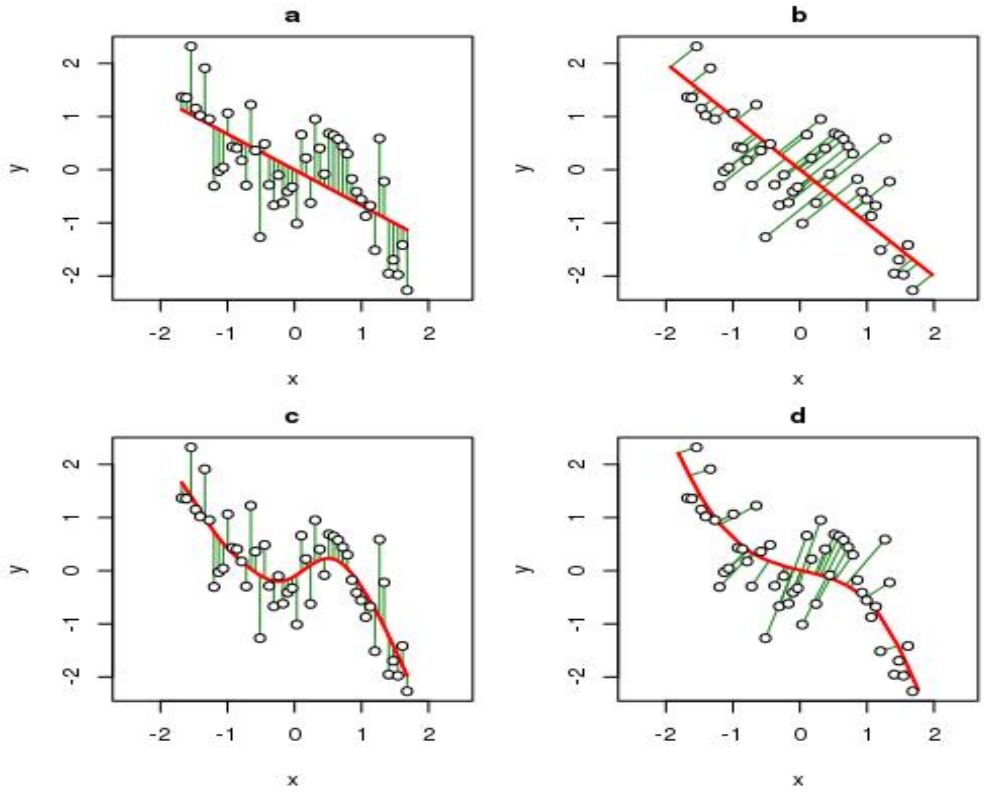


FIGURA 4 - ALGUMAS FORMAS DE AJUSTE DE DADOS
FONTE: HASTIE e STUETZLE(1989)

$$t_f(\underline{x}) = \sup\{t : \|\underline{x} - f(t)\| = inf_{\mu}\|\underline{x} - f(\mu)\|\}, \quad (81)$$

Onde $\underline{x}$ é uma observação arbitrária de X e μ é uma variável auxiliar definida em $\mathbb{R}$. O índice de projeção $t_f(\underline{x})$ é o valor de t para o qual a CP $f(t)$ está mais próxima de $\underline{x}$. Segundo Tarpey e Flury (1996) a esperança condicional $E[X \setminus t_f(\underline{x}) = t]$ é uma aproximação melhor que o erro quadrático médio. A figura 5 mostra as projeções ortogonais das observações sobre a curva principal.

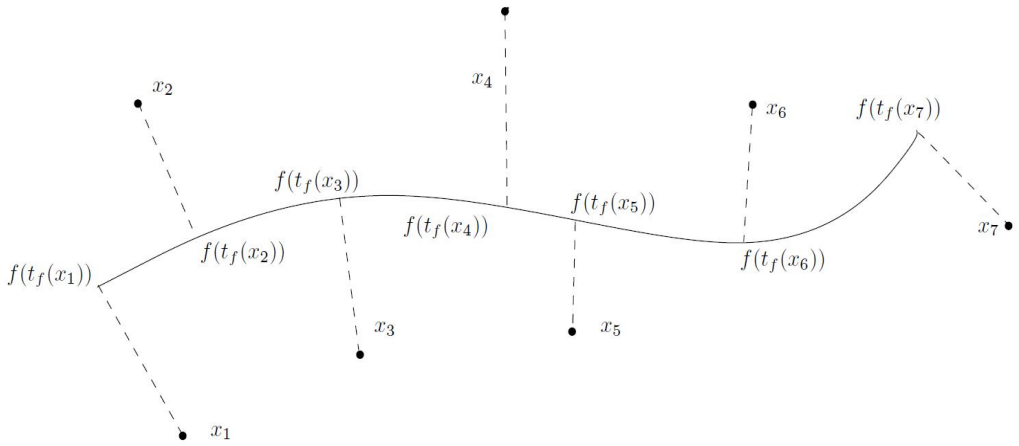


FIGURA 5 - PONTOS DE PROJEÇÃO SOBRE A CURVA $f(t)$
FONTE: O autor (2014)

O quadrado da distância euclidiana de f e $\underline{x}$ é distância ao quadrado de $\underline{x}$ ao ponto de projeção de f. Isto é:

$$\Delta(\underline{x}, f) = inf_{a \le t \le b} \|\underline{x} - f(t)\|^2 = \left\|\underline{x} - f\left(t_f(\underline{x})\right)\right\|^2. \quad (82)$$

O comprimento de uma curva f sobre um intervalo $[\alpha, \beta] \subset$ [a, b], denotado por $l(f, \alpha, \beta)$ é definido como:

$$l(f, \alpha, \beta) = sup \sum_{i=1}^{N} \|f(t_i) - f(t_{i-1})\|, \quad (83)$$

Onde o supremo é tomado sobre todas as partições finitas de [a, b] com pontos $\alpha = t_0 \le t_1 \le \cdots \le t_N = \beta$, N ≥ 1.

2.3.2.1 Propriedade geométrica para curvas

Uma importante propriedade para as componentes principais é que são pontos críticos da função distância. A primeira componente principal é a direção que minimiza a função distância entre todas as possíveis direções. Seja $d(x, f)$ a distância euclidiana de $\underline{x}$ ao ponto de projeção de f, tal que $d(x, f) = \left\|\underline{x} - f(t_f(\underline{x}))\right\|$. A função distância para a curva f é definida como a esperança do quadrado da distância euclidiana entre X e f, ou seja,

$$\Delta(h, f) = E[\left\|X - f(t_f(X))\right\|^2 \quad (84)$$

Considerando que a distribuição de X não é conhecida, mas o conjunto de dados $X_n = \{\underline{x}_1, \dots, \underline{x}_n\}$ seja independente e identicamente distribuído (i.i.d.), pode-se utilizar a função distância estimada para a curva f denominada função distância empírica, definida como:

$$\Delta_n(f) = \frac{1}{n}\sum_{i=1}^{n} \Delta(\underline{x}, f). \tag{85}$$

Considere curvas na forma:

$$s(t) = t\underline{u} + \underline{c} \tag{86}$$

Onde $\underline{c} \in \mathbb{R}^p$, e $\underline{u}$ um vetor unitário. Então, a distância ao quadrado do vetor $\underline{x}$ e a linha *s* é:

$$\Delta(\underline{x}, s) = inf\left\|\underline{x} - s(t)\right\|^2 \tag{87}$$

$$= inf\left\|\underline{x} - (t\underline{u} + \underline{c})\right\|^2 \tag{88}$$

$$= \left\|\underline{x} - \underline{c}\right\|^2 + inf_{t\in\mathbb{R}}\left(t^2 - 2t(\underline{x} - \underline{c})^T\underline{u}\right) \tag{89}$$

$$= \left\|\underline{x} - \underline{c}\right\|^2 - ((\underline{x} - \underline{c})^T\underline{u})^2 \tag{90}$$

A projeção do vetor $\underline{x}$ sobre a reta s é $\underline{c} + ((\underline{x} - \underline{c})^T\underline{u})\underline{u}$. Se $S(t) = t\underline{u} + \underline{c}$ é um segmento definido no intervalo [a, b], a distância de $\underline{x}$ ao segmento depende do índice de projeção, conforme mostra a figura 6. Se $t_s(\underline{x}) = a$ ou $t_s(\underline{x}) = b$, $\Delta(\underline{x}, s) = \left\|\underline{x} - \underline{c}\right\|^2$, onde $\underline{c}$ é um dos extremos do segmento. Caso contrário, utiliza-se a expressão (90).

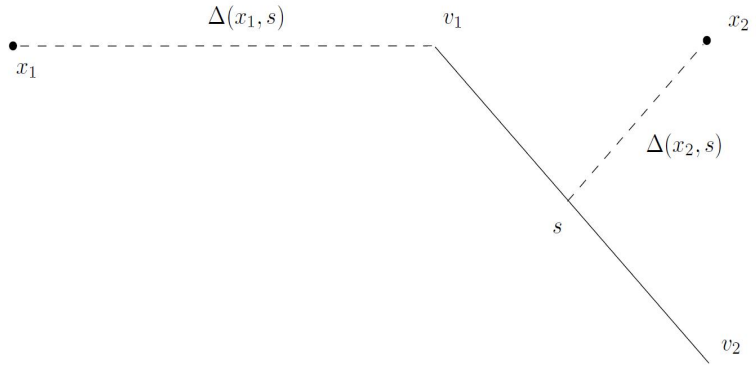


FIGURA 6 - DISTÂNCIA DE UM PONTO A UM SEGMENTO
FONTE: KÉGL *et al.* (2000)

O procedimento para encontrar a curva principal começa com a primeira componente principal linear (PCA) traçada sobre os dados. Primeiramente, os dados são projetados (ortogonais) sobre a PCA e então todos os pontos, dentro de um intervalo, são usados para calcular a média. As médias de cada subintervalo são usadas para encontrar a primeira aproximação para a curva principal (através de uma *spline*, por exemplo). O processo é iterativo, isto é, projetam-se os pontos sobre a curva, calcula-se a média de cada subintervalo e a nova aproximação da curva é obtida. Esse processo é repetido até a convergência desejada, como mostra as figuras 7 e 8.

Definições e métodos alternativos para estimar curvas principais foram apresentados após o trabalho inovador de Hastie e Stuetzle (1989). Kramer (1991) utilizou redes neurais para obter componentes principais não lineares (*NLPCA – Nonlinear principal components*) e Verbeek, Vlassis e Kröse (2002) propôs um algoritmo denominado *k*-segmentos, que é um método incremental para determinar as curvas principais e segundo os autores, o algoritmo é superior a outras técnicas para o cálculo de curvas principais.

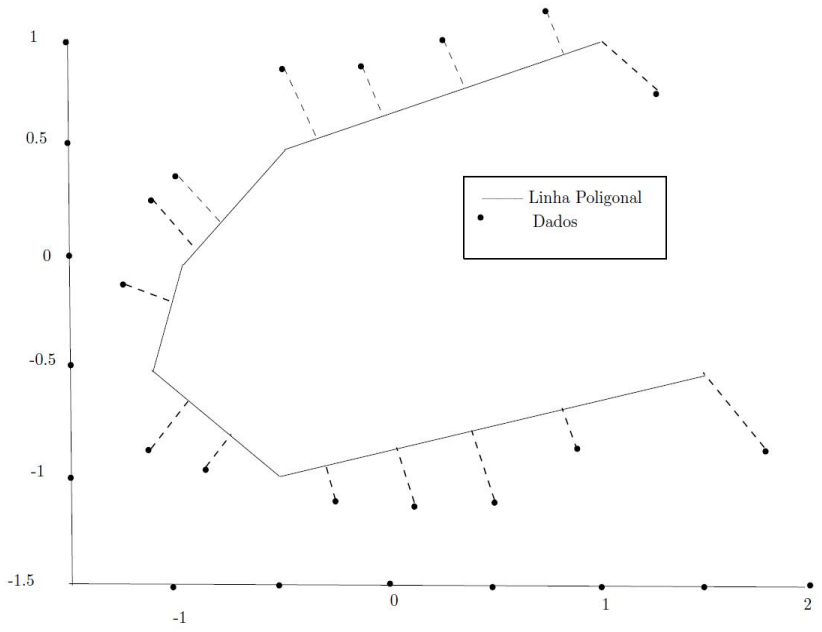


FIGURA 7 - ILUSTRAÇÃO DE PROJEÇÃO DOS PONTOS SOBRE A LINHA POLIGONAL
FONTE: KÉGL *et al*. (2000)

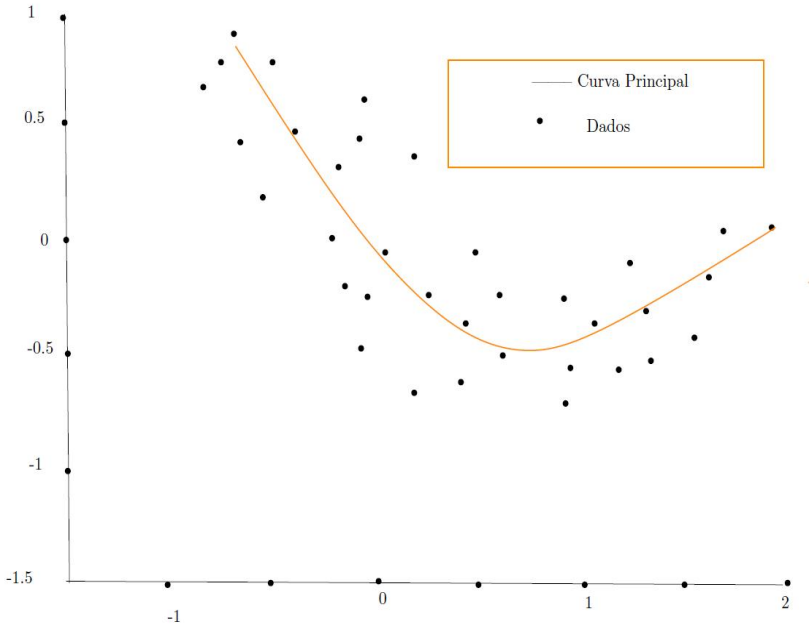


FIGURA 8 - ILUSTRAÇÃO DE CP DETERMINADA POR CONJUNTO DE DADOS
FONTE: KÉGL *et al*. (2000)

2.3.2.2 Algortimo *k*-segmentos

Um número determinado de segmentos é adicionado um por um. Esses segmentos são ligados, formando uma linha poligonal, que é suavizada para formar a curva principal. A cada iteração é inserido um novo segmento para formar a linha poligonal e a curva é novamente otimizada. Para encontrar a linha poligonal, são utilizados vários algoritmos: inicia-se com o algoritmo *k*-médias para encontrar o subconjunto V_1, que irá conter o primeiro segmento. É traçada a linha em V_1 (algoritmo *k*-linhas). Na sequência, o algoritmo *k*-linhas é adaptado para determinar o segmento pertencente à região V_1. Repete-se o procedimento para determinar as *k-1* demais regiões de Voronoi, sempre ligando os segmentos para formação da linha poligonal. A seguir serão abordadas as etapas do algoritmo.

2.3.2.3 Do algoritmo *k*-médias para *k*-linhas.

Uma linha é definida como $s(\lambda) = \underline{c} + u\lambda,\ \lambda \in \mathbb{R}$ e a distância euclidiana de uma observação amostral $\underline{x}$ à linha é definida por:

$$d(\underline{x}, s) = inf_{t\in\mathbb{R}}\|s(\lambda) - \underline{x}\|. \quad (91)$$

$V_1, V_2., \ldots, V_k$ são subconjuntos, denominados as regiões de Voronoi, de forma que $V_i = \{\underline{x} \in X_n \setminus i = argmin_j\, d(\underline{x}, s_j)\}$, isto é, essas regiões são polígonos cujo interior é constituído por todos os pontos do plano e cuja distância é a mais próxima de um ponto em particular (centroide), que contém todos os pontos mais próximos da i-ésima reta, conforme é mostrado na figura 9.

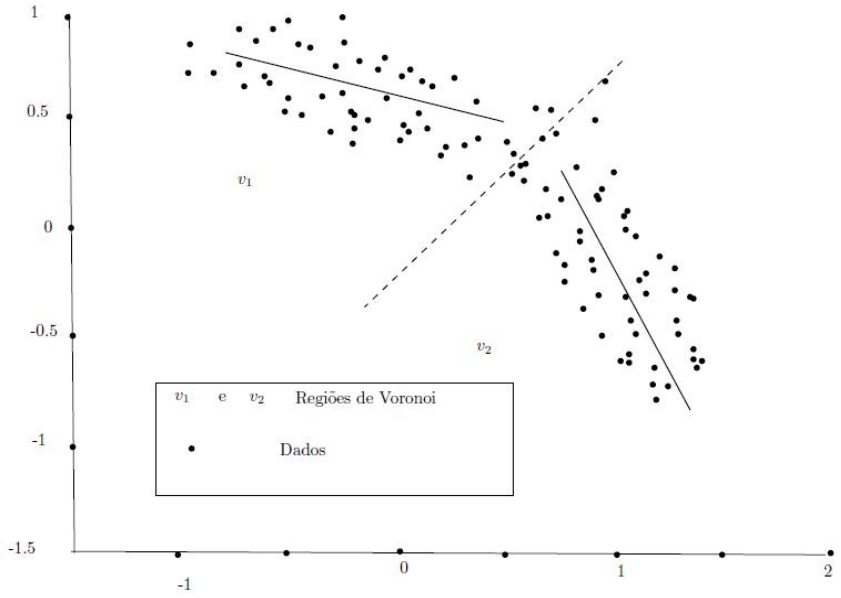


FIGURA 9 - REGIÕES DE VORONOI. CONJUNTO DIVIDIDO EM DUAS REGIÕES v_1 E v_2
FONTE: VERBEEK, VLASSIS e KRÖSE (2002)

Para encontrar $s_i, i = 1, \ldots, k,$ deve-se minimizar o total do quadrado das distâncias:

$$\sum_{i=1}^{k} \sum_{x \in V_i} d(s_i, \underline{x})^2. \quad (92)$$

Para encontrar as retas, que são ótimos locais da expressão (92), faz-se uma pequena modificação no algoritmo *k*-médias: inicia-se o algoritmo com uma orientação aleatória e são determinadas as *k*-linhas. O próximo passo alterna-se entre as duas etapas seguintes, até a convergência:

1. Determine as regiões de Voronoi;
2. Substitua as linhas com a primeira componente principal linear para cada região de Voronoi.

2.3.2.4 Do algoritmo *k*-linhas para *k*-segmentos

O comprimento de cada linha é ajustado para incluir todas as projeções ortogonais de todos os pontos na região de Voronoi. Cada linha é transformada em segmento: 'corta-se' a linha em $(3/2)\sigma$ do centroide de V_i, onde σ^2 é a variância dos pontos de projeção ortogonal sobre a primeira componente principal. Testa-se a propriedade de convergência a partir da expressão (92). Se a distância é menor, então este é o novo segmento da região V_i. Se não, usa-se o segmento que inclui todas as projeções para a primeira componente principal, de modo a garantir a redução da distância, conforme a expressão (92). Se a projeção de $\underline{x}$ muda de segmento, é repetido o processo de agrupamento (*k*-médias).

2.3.2.5 Linha poligonal

A etapa seguinte consiste em ligar os segmentos gerados em cada região de Voronoi, para formar a linha poligonal. A solução é a construção de um grafo hamiltoniano com caminho de custo total mínimo. Considere um grafo conexo *G(V, E)*, onde *V* é o conjunto de vértices, com 2*k* (o dobro do número de segmentos) elementos e o conjunto *E* o conjunto de arestas, com *k* segmentos, e define-se o conjunto $A \subset$ E de todas as arestas geradas. A sequência de vértices $v_0, v_2, \ldots, v_{m-2}, v_{m-1}, v_m$ com vértices distintos, é definida como caminho do grafo *G(V, E)*. Um caminho aberto ($v_0 \neq v_m$) é dito ser hamiltoniano (HP) se 'passa' em

cada vértice do grafo exatamente uma única vez. Seja $P \subset E$ um caminho hamiltoniano, o objetivo é minimizar o custo total do caminho *P*, com a restrição $A \subset P \subset E$. O custo para o caminho *P* é definido como:

$$l(P) + \lambda a(P), \text{ com } 0 \leq \lambda \in \mathbb{R} \quad (93)$$

Onde: $l(P)$ é o comprimento do caminho hamiltoniano, definido como a soma do comprimento das arestas de *P*. O comprimento da aresta $e = (v_i, v_j)$ é determinado por $l(e) = \|v_i - v_j\|$. O termo $a(P)$ é um termo de penalidade igual a soma dos ângulos formados com a ligação entre segmentos, conforme figura 10. O valor do parâmetro de controle λ determina a preferência com caminhos curtos e que não se cruzem. Parâmetro λ pequeno dá preferência para caminhos com curvas suaves.

O ângulo de penalidade de $(v_i,\ v_j)$ é a soma entre os ângulos de ligação dos vértices adjacentes de dois segmentos, isto é, $\alpha + \beta$.

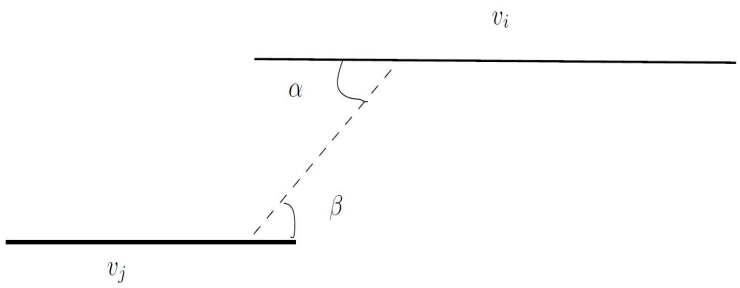


FIGURA 10 - CONEXÃO DE DOIS SUBCAMINHOS
FONTE: VERBEEK, VLASSIS e KRÖSE (2002)

O algoritmo que segue é utilizado para construção do grafo hamiltoniano.

1. Comece com *k* subgrafos hamiltonianos;
2. Enquanto existe pelo menos dois subgrafos hamiltonianos, ligue dois subgrafos de modo a minimizar o caminho *P*.

2.3.2.6 Função objetivo

Uma linha poligonal é obtida após a inserção dos *k* segmentos nas regiões de Voronoi. Agora o objetivo é, dentre todas as linhas poligonais encontradas, escolher a linha que maximiza a aproximação da log-verossimilhança dos dados selecionados. Essa aproximação é dada por:

$$n.\log(l) + \sum_{i=1}^{k}\sum_{x\in V_i} d\left(s_i, \underline{x}\right)^2/(2\sigma^2) \quad (94)$$

Onde l é o comprimento total da linha poligonal, d é à distância do ponto $\underline{x}$ à reta s_i e σ^2 é a variância dos pontos de projeção ortogonal sobre a primeira componente principal.

A seguir é apresentado o algoritmo, de modo resumido, para obtenção da linha poligonal e da CP:

1. Início: Determine a região V_1 e a linha s_1. Calcule a primeira PCA, com comprimento no intervalo [-3σ/2, 3σ/2] do centroide de V_1, onde σ^2 é a variância dos pontos de projeção ortogonal sobre a primeira componente principal. Faça i = 2.
2. Adicione novo segmento: Seja $\underline{x_i}$ o ponto de inserção ou o segmento de comprimento zero em V_i, onde V_i é a i-ésima região de Voronoi. Seja s_i a linha que minimiza a distância de s_i a $\underline{x}\in V_i, \sum_{i=1}^{k}\sum_{\underline{x}\in V_i} d\left(s_i, \underline{x}\right)^2$, troque a linha pela primeira PCA e então insira um segmento ao longo da PCA, de comprimento ±3σ/2 da média.
3. Construção e otimização: A partir da função custo $l(P) + \lambda a(P)$, construa o caminho hamiltoniano (linha poligonal) e em seguida otimize o caminho com a função objetivo $n.\log(l) + \sum_{i=1}^{k}\sum_{x\in V_i} d\left(s_i, \underline{x}\right)^2/(2\sigma^2)$.
4. Repetir 2-3.

Esse algoritmo é implementado no programa desenvolvido por Verbeek, Vlassis e Kröse (2002), para o *software* MatLab e tem o seguinte formato de entrada:

[$edges, vertices, of, y$]= $k\_seg\_soft(X, k\_max, alpha, lambda, INT\_PLOT)$

Onde:

X: Matriz de dados amostrais de ordem *n* x *p*, com *p* variáveis e *n* observações;

k_max: Número máximo de segmentos para formar a linha poligonal. O algoritmo insere os segmentos até atingir a convergência ou até o valor o número de segmentos pré-definidos;

alpha: Parâmetro de alisamento da CP, $alpha = 2.(p-1).\sigma^2$;

lambda: O parâmetro de controle λ é para ter a preferência com caminhos curtos ou que não se cruzem. Parâmetro λ pequeno dá preferência para caminhos com curvas suaves. O objetivo principal é obter curvas suaves;

INT_PLOT: Se *int_plot* = 1, o programa desenha o gráfico com os segmentos.

Variáveis de Saída:

edges: é uma matriz de ordem *2k* x *2k,* a qual contém o tipo de conexão entre os vértices. Seja a_{ij} um elemento da matriz *edges, i* e *j* são vértices adjacentes, os valores de a_{ij} são:

0 – não há conexão entre os vértices *i* e *j*;

1 – *i* e *j* são os vértices de uma aresta de ligação. São vértices de arestas criadas na construção da linha poligonal;

2 – *i* e *j* são vértices do segmento de ligação obtido nas regiões de Voronoi.

Por exemplo, dada a matriz *edges*,

$$edges = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 2 & 0 \end{bmatrix}$$

A linha poligonal é formada pelas arestas de vértices (1, 2), (4, 3) e (5, 6) e as arestas de ligação são (2, 3) e (4, 5).

vertices: é uma matriz de ordem 2k x p. Contém as coordenadas dos vértices dos segmentos que compõem a curva principal;

of: contém avaliações da função objetivo (terceira coluna), a distância média ao quadrado da curva (primeira coluna) e log do comprimento da curva (*log(l)*) (segunda coluna);

y : contém coluna dos dados projetados sobre a curva (colunas 2 até a última) mais os índices de projeção (primeira coluna).

2.3.2.7 Exemplo numérico do algoritmo *k*-segmentos

A seguir são apresentadas duas simulações utilizando o algoritmo *k*-segmentos. Para o primeiro exemplo, foi definida a seguinte função:

$$y = x^2, \text{ com } x \in [-3, 3]. \quad (95)$$

A figura 11 apresenta a curva principal e a linha poligonal obtida sem ruído branco. A matriz de entrada possui duas colunas: a primeira contém os valores de x e a segunda os valores de y. É fácil ver que o ajuste feito pelo algoritmo que gerou a parábola foi igual ao gráfico da função, mostrando a eficiência do algoritmo. A seguir, os comandos do MatLab para gerar a CP.

```
>> x= -3:0.01:3;
>> y= x.^2;
>> xy= [x; y];
>> [vertices,edges,c,d]=k_seg_soft(xy',2,1,1,1)
```

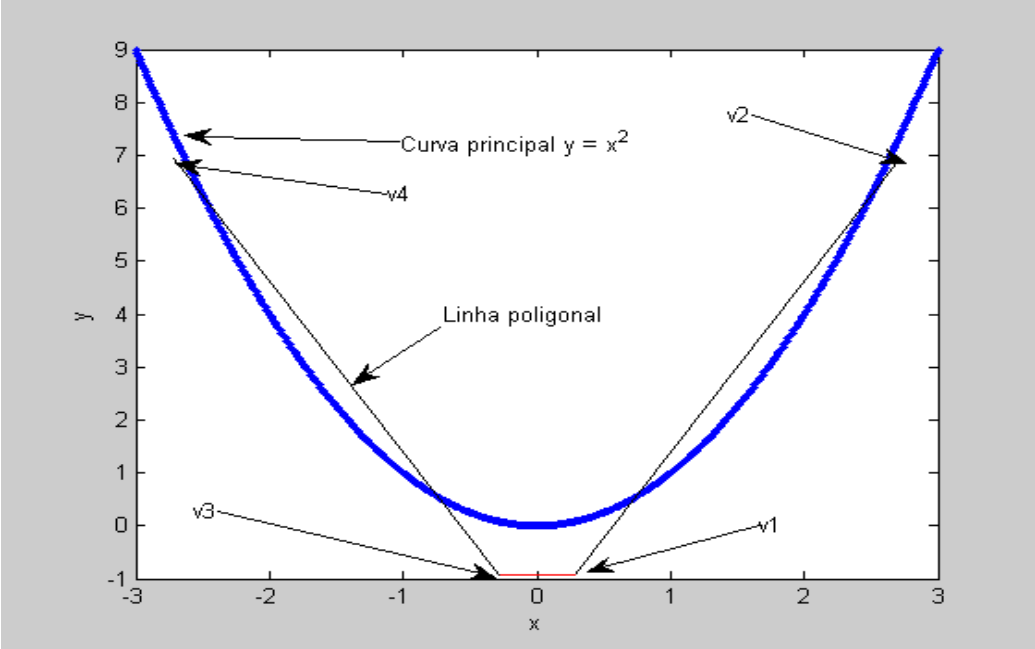


FIGURA 11 - CURVA PRINCIPAL DA CP
FONTE: O autor (2014)

O número máximo de segmentos é 2 (*k*-max = 2), lambda igual 1, com o objetivo de encurtar ao máximo a ligação entre os segmentos e tornar a curva mais suave. A matriz *edges* apresenta os segmentos que formam a linha poligonal. Na figura 11, as duas arestas geradas tem como vértices adjacentes v_1 e v_2, e v_3 e v_4, e os vértices da aresta de ligação são v_3 e v_4. As coordenadas de cada vértice são dadas pela matriz *vertices*.

$$edges = \begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$

$$vertices = \begin{bmatrix} v1 & v2 & v3 & v4 \\ 0,283 & 2,711 & -0,283 & -2,711 \\ -0,933 & 6,953 & -0,933 & 6,953 \end{bmatrix}$$

Considerando a matriz *edges* e os elementos acima da diagonal principal, o único elemento com valor 1 (um), *edges* (1, 3), indica que o algoritmo construiu a linha poligonal com apenas um segmento de ligação e este segmento tem como coordenadas os vértices v1 e v3, que são: (0,283; -0,933) e (-0,283; -0,933).

No segundo exemplo, foi definida a função $y = sen(x) + \varepsilon$, cuja componente ε é uma variável independente (ruído branco) distribuída uniformemente no intervalo [-0,2; 0,2] e *x* é uma variável definida sobre o intervalo [0; 2π], com intervalos de 0,1, isto é, *x* = [0; 0,1; 0,2; ... ; 2π]. A matriz gerada pelos valores de *x* e *y* é de ordem 63x2. A figura 14 apresenta a linha poligonal com 7 segmentos, sendo 3 de ligação. A seguir, os comandos do MatLab para gerar a linha poligonal e a curva principal.

```
>> xsen=0:0.1:2*pi; % gera o vetor de pontos para x
>> ysen=sin(xsen); % gera o vetor de pontos para y
>>rdsen=-0.2+(0.2+0.2).*rand(63,1);  % ruído ou erro
>> ysenrd=ysen+rdsen'; % adicionado ruído à função seno.
>> xysenrd=[xsen;ysenrd]; % constroi a matrix com as colunas das variáveis x e y
>> [vertices,edges,c,d]=k_seg_soft(xysenrd',4,1,1,1) % comando para gerar a curva
% principal com 4 segmentos.
```

A figura 12 ilustra a linha poligonal obtida com *k*_max = 4, com três segmentos de ligação. As coordenadas dos vértices dos extremos de cada segmento são dados pela matriz *vertices* e a matriz *edges* informa a ordem dos vértices para cada segmento. A matriz com as coordenadas dos vértices, gerada pelo algoritmo *k*-segmento, é dada por:

$$vertices = \begin{bmatrix} v1 & v2 & v3 & v4 & v5 & v6 & v7 & v8 \\ 4,46 & 2,15 & 1,42 & 0,069 & 6,126 & 4,783 & 1,971 & 1,568 \\ -1,18 & 0,85 & 1,19 & 0,063 & -0,180 & -1,198 & 0,869 & 0,851 \end{bmatrix}$$

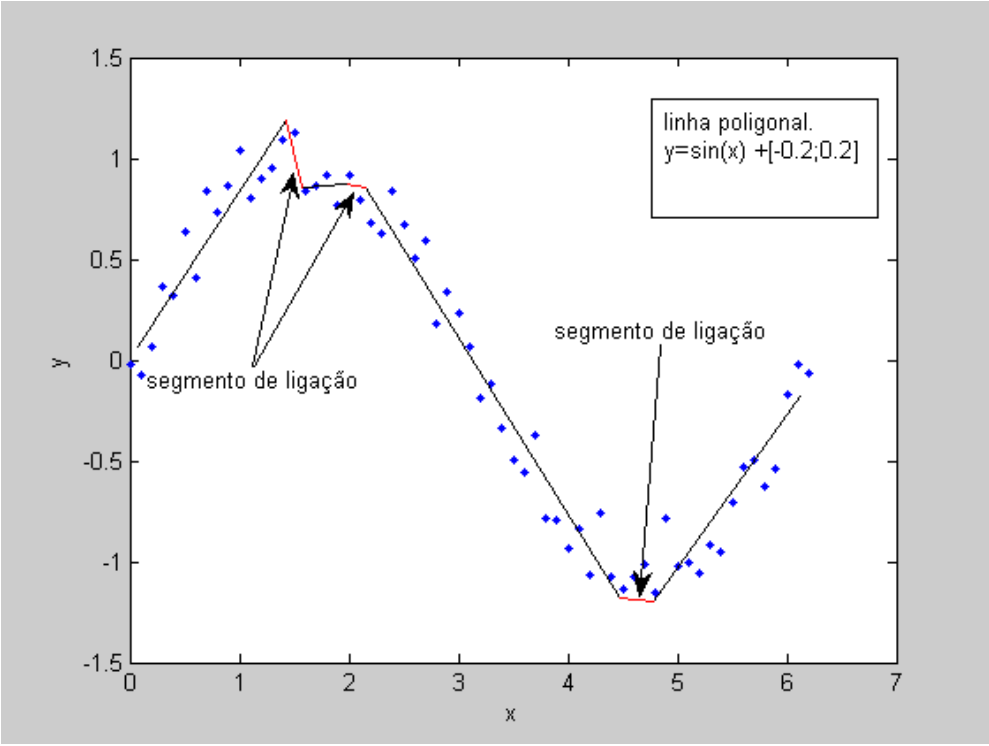


FIGURA 12 - LINHA POLIGONAL DA FUNÇÃO SENO
FONTE: O autor (2014)

E a matriz com a ordem de ligação dos vértices que formam o segmento é dada por:

$$edges = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 2 & 0 \end{bmatrix}$$

Observando a matriz *edges,* os elementos *edges* (1,6), *edges* (2, 7) e *edges* (3, 8), da parte superior a diagonal principal, têm valor 1 e indicam os vértices dos extremos dos segmentos de ligação da linha poligonal. Considerando os segmentos de ligação da esquerda para a direita da linha poligonal, apresentada na figura 13, têm-se os seguintes vértices:

(v1; v6) ⇨ (4,46; -1,18) e (4,783; -1,198)
(v2; v7) ⇨ (2,15; 0,85) e (1,917; 0,869)
(v3: v8) ⇨ (1,42; 1,19) e (1,568; 0,851)

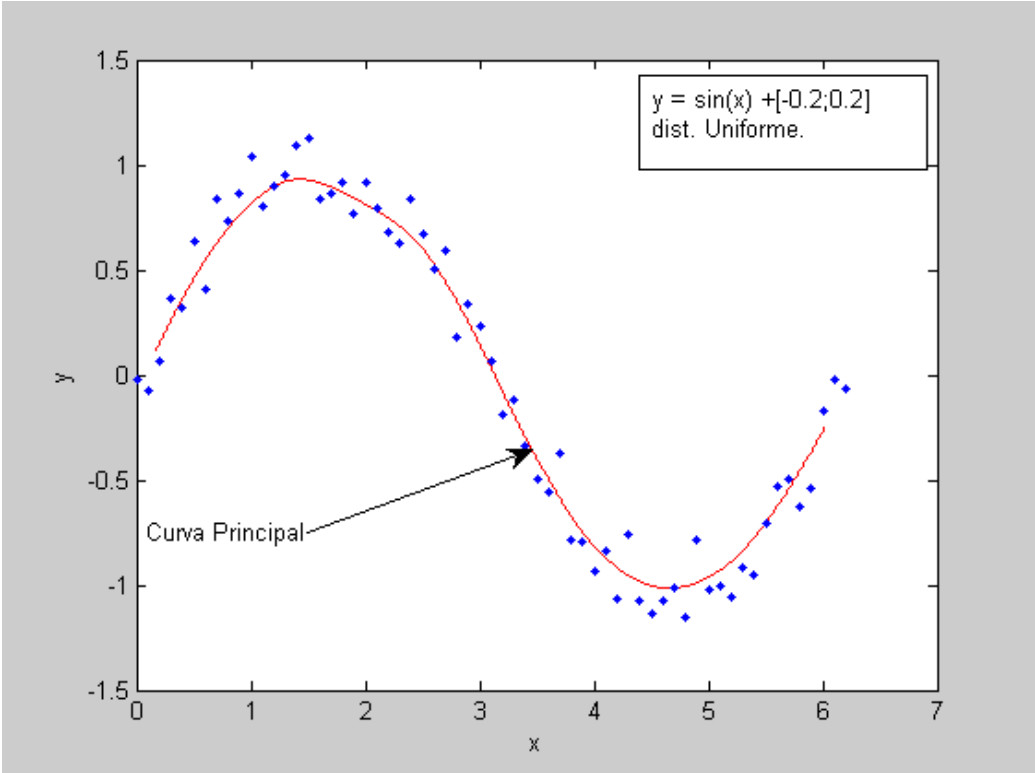


FIGURA 13 - CURVA PRINCIPAL DA FUNÇÃO SENO
FONTE: O autor (2014)

A curva gerada pelo algoritmo sugere a forma de uma senoide, resultado já esperado pelo ajuste feito pelo algoritmo. Também está em conformidade com a definição de curvas principais, de modo que a senoide (curva principal) gerada é uma linha que ‘passa’ pelo meio dos pontos.

3 MATERIAL E MÉTODOS

O procedimento de classificação consiste na escolha de um método formal capaz de decidir, com base nas observações experimentais, a que grupo ou população um determinado objeto pertence. A eficiência na classificação é importante para qualquer método de classificação aplicado. O algoritmo *k*-segmentos, neste trabalho, é utilizado com o objetivo de melhorar a eficiência de classificação das técnicas FDA e LFDA. Convém destacar que, segundo Hand (2006) a FDA é uma ótima ferramenta para classificação supervisionada com muitas aplicações, devido a sua simplicidade, robustez e eficiência preditiva. Também, segundo Lim e Shih (2000), nenhum procedimento formal para este fim é absolutamente mais eficiente e que a técnica FDA linear apresenta, em geral, melhor eficiência que outras técnicas de classificação, inclusive em relação à análise discriminante quadrática.

Neste capítulo, é inicialmente feita a introdução sobre o tema abordado no presente trabalho, em seguida é apresentada a técnica para a classificação baseada na análise discriminante. Na subseção 3.2.1, é descrita a técnica na forma de algoritmo, denominado algoritmo *k*-segmentos. E finalmente na última subseção são apresentados os conjuntos utilizados para avaliação do desempenho da técnica para classificação.

3.1 MÉTODO PROPOSTO

O método desenvolvido neste trabalho faz uma modificação nos métodos LFDA e FDA, com o objetivo de tornar a classificação mais eficiente. Esta modificação é efetuada após o cálculo dos escorres discriminantes, no instante em que são determinadas as distâncias do objeto a ser classificado em uma das classes previamente conhecida. A FDA utiliza a distância euclidiana aos centroides de cada classe, o método proposto substitui os centroides pelas linhas poligonais geradas pelo algoritmo *k*-segmento que também gera as curvas principais através de linhas poligonais.

3.2 CLASSIFICAÇÃO BASEADA EM CURVAS PRINCIPAIS

O procedimento de classificação consiste em alocar um novo objeto a uma classe dentre *k* classes conhecidas *a priori*. Parte-se do conhecimento de que os *n* indivíduos observados pertencem a diversos subgrupos e procura-se determinar funções das *p* variáveis observadas que melhor permitam distinguir ou discriminar entre subgrupos ou classes. A figura 14 (fluxograma adiante) mostra os passos para a análise discriminante, tanto na FDA como na LFDA.

Primeiro, são verificados os pressupostos para o uso da análise discriminante no conjunto em questão. No passo seguinte são determinadas as *k-1* combinações lineares (funções discriminantes) que melhor discriminam as classes. Os coeficientes das funções discriminantes são denominados escores discriminantes (os escores são determinados através de operações com as matrizes de covariâncias entre classes (S_B) e a matriz de covariâncias dentro das classes (S_W), pela função *J(w)*, descrita na subseção 2.1.2). No passo seguinte é feita a análise para verificar se as funções discriminantes são estatisticamente significativas. Cada observação a ser classificada deve ser estimada para cada uma das *k-1* funções discriminantes, nas quais são obtidos os escores de cada função discriminante para o novo valor observado. Pelo método dos centroides, são calculadas as médias dos escores discriminantes, denominadas de centroides, das *k* classes. Em seguida são calculadas as distâncias do vetor a ser classificado com os centroides de cada classe. O novo objeto é classificado na classe que tem menor distância euclidiana. A figura 14 mostra o fluxograma para a classificação do vetor $\underline{x}$ para um conjunto com 3 classes pela FDA e LFDA. Para cada classe é determinada a média do escore (centroide) e em seguida são calculadas as distâncias aos centroides d1, d2 e d3. Por exemplo, se d2 é a menor distância dentre as três, $\underline{x}$ é classificado na classe 2, porque d2 é a menor distância de $\underline{x}$ ao centroide da classe 2 e de qualquer outra classe.

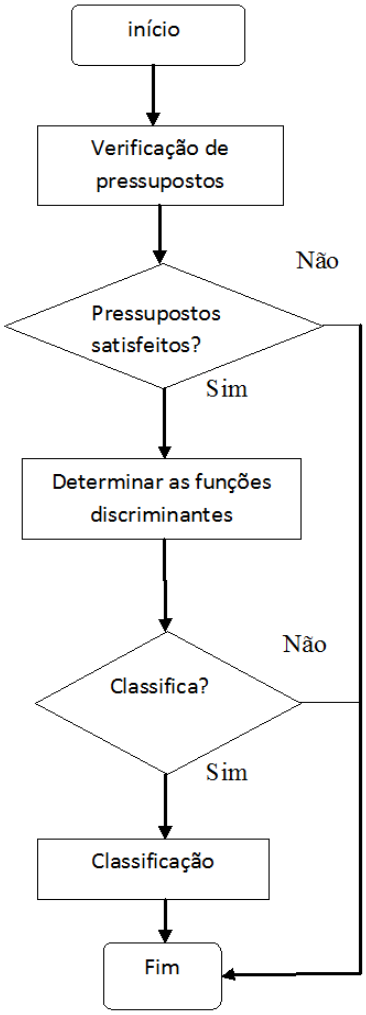


FIGURA 14 - FLUXOGRAMA PARA A ANÁLISE DISCRIMINANTE
FONTE: O autor (2014)

O método proposto neste trabalho consiste na aplicação das curvas principais (CPs) na classificação de dados amostrais. O que se procura é uma metodologia alternativa a dos centroides que melhore a eficiência na classificação, na análise discriminante.

O classificador baseado em curvas principais é aplicável a conjuntos com 3 ou mais classes. A FDA e LFDA determinam *k-1* funções discriminantes, e consequentemente para conjuntos com apenas duas classes é gerada apenas uma função discriminante, isto é, apenas uma matriz coluna 1 x n com os escores

discriminantes. Como a técnica *k*-segmentos é uma metodologia multivariada, não é viável aplicá-la a apenas uma função discriminante.

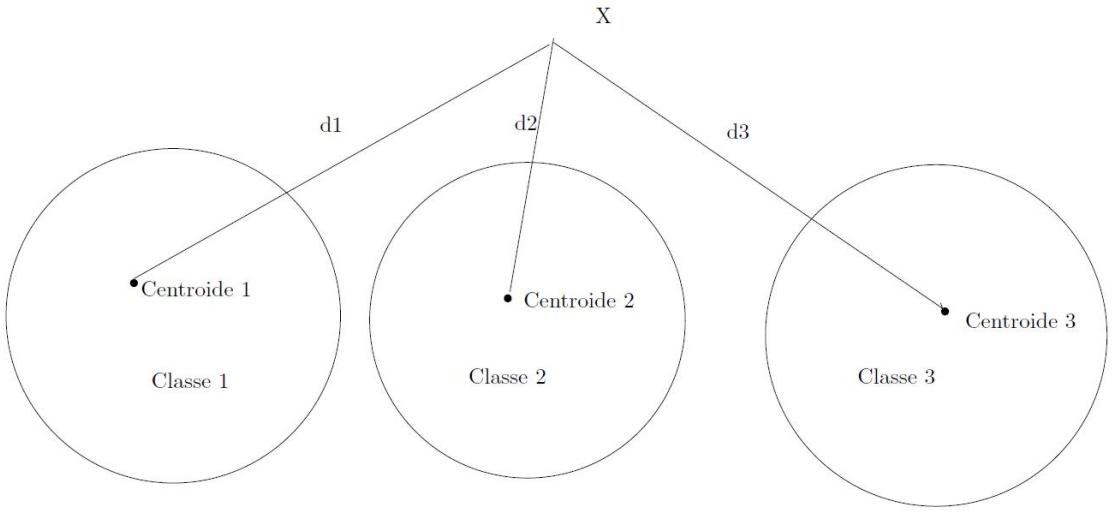


FIGURA 15 - DIAGRAMA PARA CLASSIFICAÇÃO DO VETOR $\underline{X}$ POR MEIO DE CENTROIDES PARA UM CONJUNTO COM 3 CLASSES
FONTE: O autor (2014)

Este classificador é incorporado em duas técnicas, já descritas na seção 2, para classificação: a análise discriminante de Fisher (FDA) e a análise discriminante local de Fisher (LFDA), também com o objetivo de comparar a eficiência das duas técnicas na classificação. Inicialmente, o procedimento para classificação é igual aos dois métodos com o uso dos centroides. Primeiramente é feita a análise dos pressupostos para o uso da FDA ou LFDA, isto é, assume-se que as matrizes de variância/covariância são homogêneas entre grupos. Se não houver essa homogeneidade, haverá uma tendência de classificar as observações no grupo de menor variância. Também neste caso, desvios menores são pouco importantes pelo que é possível uma decisão para prosseguir na análise discriminante, apesar da violação do pressuposto. A FDA é uma técnica robusta à violação desses pressupostos desde que a dimensão do menor grupo seja superior ao número de variáveis em estudo (HAIR *et al.*, 2005). A violação desse pressuposto aumenta a probabilidade de classificar observações no grupo que possuir maior dispersão.

Em seguida, são obtidos os escores discriminantes do conjunto com as *k* classes. A aplicação do algoritmo *k*-segmentos inicia com a separação do conjunto de escores nas respectivas classes, de modo que cada classe contenha os escores correspondentes aos elementos originais de sua respectiva classificação. O algoritmo é aplicado aos escores de cada classe, obtendo-se a linha poligonal, formada por vários segmentos, e a curva principal de cada classe (a curva principal não é utilizada na classificação, apenas para análise final). Para a classificação de uma nova observação, primeiramente é transformada esta observação em escore e em seguida é feita a projeção ortogonal do vetor escore sobre cada segmento da linha poligonal. A projeção determina em cada segmento o ponto de projeção. Calcula-se a distância de cada ponto de projeção ao vetor escore. O elemento é classificado na classe que contém o ponto de projeção com a menor distância do vetor escore a ser classificado. A figura 16 mostra o diagrama para a classificação do vetor $\underline{x}$ para um conjunto com 3 classes.

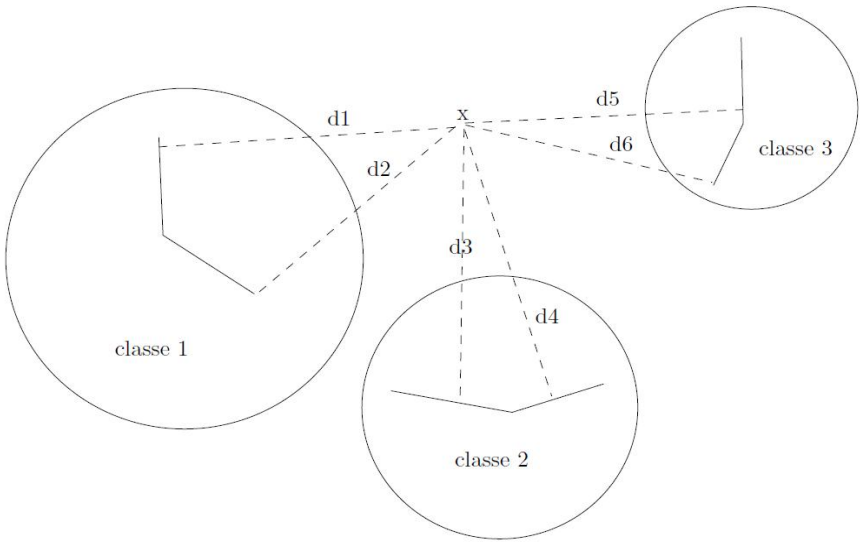


FIGURA 16 - DIAGRAMA PARA A CLASSIFICAÇÃO POR MEIO DO ALGORITMO K-SEGMENTOS PARA UM CONJUNTO COM 3 CLASSES E LINHAS POLIGONAIS COM 2 SEGMENTOS
FONTE: O autor (2014)

Para cada classe foi obtida a linha poligonal com dois segmentos. Para cada segmento é determinado o ponto de projeção ortogonal. Em seguida são calculadas

as distâncias d1, d2, d3, d4, d5 e d6. A distância di (i = 1, …6) é a distância do vetor $\underline{x}$ ao seu ponto de projeção ortogonal sobre o segmento. Se por exemplo, d3 é a menor distância dentre as seis, $\underline{x}$ é classificado na classe 2, pois d3 é a distância de x a um segmento pertencente à classe 2.

3.2.1 Algoritmo

A fim de avaliar a eficiência da técnica, foi desenvolvido um programa em MatLab para determinar as distâncias da observação amostral candidata à classificação. Foi utilizado um programa para gerar os escores discriminantes, outro para gerar as linhas poligonais e outro para calcular a distância do vetor a ser classificado a cada segmento. No passo seguinte, utilizou-e também um programa para a classificação (neste passo pode ser usado o Excel). E, finalmente, um programa para gerar as matrizes de confusão para avaliação dos resultados obtidos.

Para gerar os escores discriminantes, foi utilizado um programa desenvolvido no *software* MatLab, denominado disc2, adaptação do programa disc1[2]. O elemento de entrada do programa disc2 é a matriz de dados X, de tamanho n x p, isto é, com n observações e p variáveis e a matriz V com os tamanhos dos grupos, na ordem: primeiro grupo, segundo grupo, etc. O s*oftware k-segment*, desenvolvido por Verbeek, Vlassis e Kröse (2002), foi utilizado para obter as linhas poligonais traçadas em cada classe. E, finalmente, foram desenvolvidos dois programas para 'chamar' os programas do MatLab, para calcular os escores discriminantes e as linhas poligonais e, em seguida, calcular também as distâncias dos centroides e das linhas poligonais, pelas duas técnicas, FDA e LFDA. A classificação pode ser feita em planilhas do Excel ou MatLab.

Outra proposta é verificar a eficiência de classificação da LFDA com a FDA, por meio das duas técnicas: Centroides e *k*-segmentos. A primeira utiliza os centroides (média dos escores discriminantes) de cada classe. A classificação de uma nova observação amostral $\underline{x}$ é feita na classe cuja distância ao centroide é a menor. Essa técnica é a mesma utilizada pela FDA, da qual a LFDA é derivada. Já na técnica que utiliza o algoritmo *k*-segmentos, inicialmente procedem-se da mesma

[2] Programa desenvolvido por Jair Mendes Marques.

forma que no método anterior. Os escores discriminantes são determinados a partir do conjunto de dados amostrais e, depois, são separados por suas respectivas classes. No passo seguinte, ao invés de calcular os centroides para cada classe, são determinadas as linhas poligonais das classes $C_1, C_2, ..., C_k$ por meio do algoritmo *k*-segmentos e, finalmente, calcula-se a distância da projeção ortogonal da observação amostral $\underline{x}$ sobre cada linha poligonal e $\underline{x}$ é atribuído à classe com a menor distância de $\underline{x}$.

O algoritmo classificador *k*-segmentos é descrito pelos seguintes passos:

Dados de entrada: o conjunto de dados amostrais X com *k* classes e o número de segmentos por classe(s).

1. Inicie com a análise discriminante linear (LFDA ou FDA). Faça os testes dos pressupostos iniciais. Se as condições iniciais são satisfeitas, obtenha a matriz dos pesos discriminantes *W*, isto é, a matriz dos coeficientes discriminantes que maximizam a razão das variâncias entre as classes e dentro delas.
2. Faça $Y = W^t X$ (matriz das funções discriminantes). Separe *Y* por classe, C_i, i = 1, …, *k*.
3. Aplique o algoritmo *k*-segmentos para cada C_i. A cada C_i, são obtidas as matrizes de nós (*edges*) e de suas coordenadas (*vertices*). Os extremos dos segmentos são dados pela matriz *vertices* e a ordem de ligação dos segmentos é dado pela matriz *edges*.
4. Determine o ponto de projeção ortogonal $\underline{x}'_j$, j = 1,..., s, onde s é o número de segmentos por classe. Projeção do vetor $\underline{x}$ sobre cada segmento de C_i. Cálculo das distâncias euclidianas de $\underline{x}'_j$ a $\underline{x}$ para cada segmento da classe C_i.
5. Classifique o vetor observado na classe com a menor distância.

A vantagem desse algoritmo é que o mesmo não utiliza apenas um valor central, pois as curvas principais são formadas por pontos autoconsistentes que também são pontos médios (HASTIE, TIBSHIRANI e FRIEDMAN, 2009). A linha gerada pelo algoritmo pode determinar um ajuste não linear aos elementos da classe, o que possibilita melhorar a eficiência de classificação. A principal desvantagem dessa técnica é computacional, pois o algoritmo *k*-segmentos utiliza vários outros algoritmos e o desenvolvimento de um *software* é de complexidade

muito superior ao cálculo dos centroides. Outra vantagem é quando o conjunto de escorres discriminantes das classes está concentrado em linha ou alguma forma de ajuste não linear, a linha poligonal é ajustada para forma do conjunto com os escores próximos a linha poligonal. O algoritmo pode não ser tão eficiente nos conjuntos em que a transformação efetuada pela análise discriminante não separa efetivamente as classes, isto é, a interseção entre classes é grande, a distribuição dos escores discriminantes tem distribuição esférica ou quando há o cruzamento das curvas principais.

3.3 CONJUNTO DE DADOS

Para a avaliação da metodologia utilizada neste trabalho e também para comparação da eficiência na classificação das técnicas FDA e LFDA, são utilizados vários conjuntos amostrais, específicos para a análise discriminante, com três ou mais classes (ou grupos). A maior parte dos conjuntos foi obtida no repositório de dados da Universidade da Califórnia em Irvine (*UCI - Machine Repository,* 2014), banco de dados para aprendizado de máquina muito utilizado no meio científico. Esse repositório de banco de dados é usado, especialmente, pela comunidade de aprendizado de máquina para análise de algoritmos. A maioria dos conjuntos são contribuições feitas por pesquisadores de aprendizado de máquina que contribuíram com o banco de dados da *UCI*, sendo sendo grande parte deles para classificação. Os conjuntos selecionados são todos com variáveis aleatórias reais. Outros conjuntos também foram utilizados e serão citados adiante. Para o conjunto *segment*, foram retiradas as variáveis 4, 5 e 6 por resultarem em matrizes singulares. A Tabela 2 relaciona todos os conjuntos trabalhados, a sua origem, o total de observações, o número de variáveis e o total de classes.

TABELA 2 - CONJUNTOS AMOSTRAIS UTILIZADOS

CONJUNTO	FONTE	Nº DE OBSERVAÇÕES	Nº DE VARIÁVEIS	Nº DE CLASSES
Wine	UCI	178	13	3
Abalone	UCI	4177	8	3
Tiroide	UCI	215	5	3
Besouro	Lubischew (1962)	74	7	3
Futebol	Rencher (2002)	90	6	3
Álcool	Tanagra *datasets*	77	6	3
Glass	UCI	214	9	6
Segment	UCI	2310	19	7
Wave	UCI	5000	21	3
Iris	UCI	150	4	3
Medical	Dugard, Todman, Staines (2009)	54	3	3
Letter	UCI	20000	17	26
Veículo	UCI	846	9	4
Balance	UCI	625	4	3

FONTE: O autor (2014)

4 RESULTADOS

4.1 TESTES

Os Resultados dos testes são apresentados a seguir. Deve-se considerar que esse método não necessita da suposição de que as diversas populações sejam normais, entretanto assume-se que as matrizes de covariâncias populacionais Σ's sejam iguais, isto é:

$$\Sigma_1 = \Sigma_2 = \ldots \ \Sigma_k = \Sigma. \quad (96)$$

Caso a homogeneidade das matrizes de covariância seja violada, haverá um aumento da probabilidade para classificar observações no grupo que possuir a maior dispersão (FÁVERO *et al., 2009*). A estatística lambda de Wilks é usada para a avaliação global do conjunto de dados, pois verifica a diferença entre os centroides das classes. Quando o valor de *p* está perto de 0 (zero), significa que se tem um classificador eficiente. Esse ponto de vista está intimamente relacionado com a MANOVA. Também, para avaliar a diferença entre os centroides é utilizado o teste de Bartlett, no qual é avaliado o *p*-valor, cujo resultado deve ser inferior a 0,05 ao nível de significância de 5%. A Tabela 3 apresenta os resultados desses testes para cada conjunto.

TABELA 2 - TESTES ESTATÍSTICOS PARA OS CONJUNTOS

CONJUNTO	FUNÇÃO	Λ DE WILK'S	BARTLETT (χ2)	*p*-VALOR
Wine	1	0,02	666,79	0
	2	0,19	276,28	0
Tiroide	1	0,12	446,49	0
	2	0,58	115,37	0
Iris	1	0,02	545,58	0
	2	0,78	35,64	0
Álcool	1	0,16	132,54	0
	2	0,75	20,99	0,0008
Glass	1	0,08	522,92	0
	2	0,43	173,59	0
	3	0,71	71,69	0

continua

CONJUNTO	FUNÇÃO	Λ DE WILK'S	BARTLETT ($\chi 2$)	*p*-VALOR
				conclusão
	4	0,87	29,72	0,0031
	5	0,94	12,15	0,0328
Wave	1	0,29	6125,44	0
	2	0,57	2827,72	0
Medical	1	0,13	102,04	0
	2	0,48	36,01	0
Abalone	1	0,63	1901,61	0
	2	0,99	38,95	0
Besouro	1	0,01	309,55	0
	2	0,20	108,65	0
Futebol	1	0,31	99,76	0,00
	2	0,90	9,27	0,10
Segment	1	0,00	20654,29	0
	2	0,00	13123,08	0
	3	0,06	6622,84	0
	4	0,24	3277,55	0
	5	0,65	1002,14	0
	6	0,89	258,25	0
Letter	1	0,00	138407,19	0
	2	0,00	107361,99	0
	3	0,02	82250,66	0
	4	0,04	64728,33	0
	5	0,09	47866,18	0
	6	0,16	36521,29	0
	7	0,26	26886,21	0
	8	0,39	19033,92	0
	9	0,54	12397,97	0
	10	0,65	8533,29	0
	11	0,77	5348,00	0
	12	0,88	2526,81	0
	13	0,93	1420,33	0
	14	0,96	716,77	0
	15	0,99	119,17	0
	16	1,00	12,15	0,2749
Veículo	1	0,0833	2071,7516	0,00
	2	0,286558	1042,34	0,00
	3	0,870022	116,12	0,00
Balance	1	0,3231	700,9954	0,00
	2	0,9993	0,3959	0,9411

FONTE: O autor(2014).

Na Tabela 3 (ver páginas 66 e 67) são apresentados os resultados do teste de capacidade discriminatória das funções, para cada um dos conjuntos trabalhados. Por exemplo, para o conjunto *wine,* tem-se que Λ = 0,02 e Λ = 0,19, para as funções 1 e 2 respectivamente, valores que indicam bons resultados para esse teste (o pior valor é 1). No teste de Bartlett, *p*-valor apresentou os valores p = 0 e p = 0 (menores que 0,05) para as duas funções discriminantes, e ao nível de significância de 5%, rejeita-se a hipótese nula para as funções 1 e 2, isto é, os centros dos grupos são significativamente diferentes. No conjunto *letter* são utilizadas 20.000 observações para desenvolver um modelo discriminante para as 26 classes que compõem esse conjunto. Apenas 15 funções discriminantes, com *p*-valor menor que 0,05, são estatisticamente significativas ao nível de significância de 5%. Para os demais conjuntos, a interpretação é semelhante a essa análise.

4.2 FDA - CENTROIDES *VERSUS* FDA *K*-SEGMENTOS

Nesta seção é feita a aplicação do algoritmo *k*-segmentos para os conjuntos já citados com a análise discriminante de Fisher (FDA). Em cada conjunto é aplicada a análise discriminante, a FDA, para obtenção da matriz dos escores discriminantes. A matriz dos escores é particionada em *k* classes, obtendo-se as submatrizes dos escores de suas respectivas classes. Em cada submatriz é aplicado o algoritmo *k*-segmentos para obtenção dos segmentos que formam a linha poligonal. Em seguida, são obtidas as distâncias do elemento a ser classificado aos segmentos.

De modo a comparar os resultados obtidos de cada uma das metodologias, centroide e *k*-segmentos, foram confrontados os valores obtidos através dos erros de classificação. Os resultados são apresentados através da matriz de confusão para cada conjunto de dados.

4.2.1 Iris

Este conjunto, devido a Fisher, é um conjunto bastante conhecido na análise discriminante (JOHNSON e WICHERN, 1998). É o conjunto utilizado por Fisher para apresentar a análise discriminante ao meio científico em seu artigo "*The use of multiple measurements in axonomic problems"* em 1936. É composto por 3 classes, com 50 amostras em cada classe e 4 variáveis. As classes são as três espécies da planta *Iris: setosa, versicolor, virginica* (FIGURA 17).

FIGURA 17 - ESPÉCIES DE IRIS
FONTE: <http://slideplayer.pl/slide/836104/>. Acesso em 15/08/2015

A matriz de confusão para cada método é apresentada na tabela 4. Os resultados mostram o bom desempenho dos dois métodos, com uma pequena vantagem para o método *k*-segmentos. Este bom desempenho pode ser explicado pelo fato da análise discriminante ter sido eficiente na separação das classes, com grupos bem definidos e com centroides distintos, cuja distância é desejável entre os mesmos, conforme é observado na figura 18.

Embora esse conjunto contenha apenas 150 observações, o tamanho da amostra e o tamanho dos grupos estão dentro do que é sugerido na análise discriminante. Segundo Hair *et al*., (2005) a análise discriminante é muito sensível à proporção entre o tamanho da amostra e o número de variáveis preditoras e sugere que uma proporção de no mínimo 20 observações para cada variável preditora.

TABELA 3 - MATRIZ DE CONFUSÃO PARA O CONJUNTO IRIS PARA FDA CENTROIDES X FDA *K*-SEGMENTOS

		FDA CENTROIDES			FDA *K*-SEGMENTOS *k=2*		
CLASSE	**TAMANHO DA CLASSE**	**CLASSE PREDITA**			**CLASSE PREDITA**		
		1	**2**	**3**	**1**	**2**	**3**
1	50	**50**	0	0	**50**	0	0
		100%	0%	0%	**100%**	0%	0%
2	50	0	**48**	2	0	**49**	1
		0%	**96%**	4%	0%	**98%**	2%
3	50	0	1	**49**	0	1	**49**
		0%	2%	**98%**	0%	2%	**98%**
Casos classificados corretamente: 98%							**98,66%**

FONTE: O autor (2015)

Através das figuras 18 e 19, é fácil ver que a análise discriminante é eficiente na separação das classes deste conjunto, com pequena interseção nas classes 2 e 3 e com os centroides no centro de cada classe. As curvas principais 'passam' pelo meio de cada classe, apresentando um ajuste eficiente da CP para cada classe.

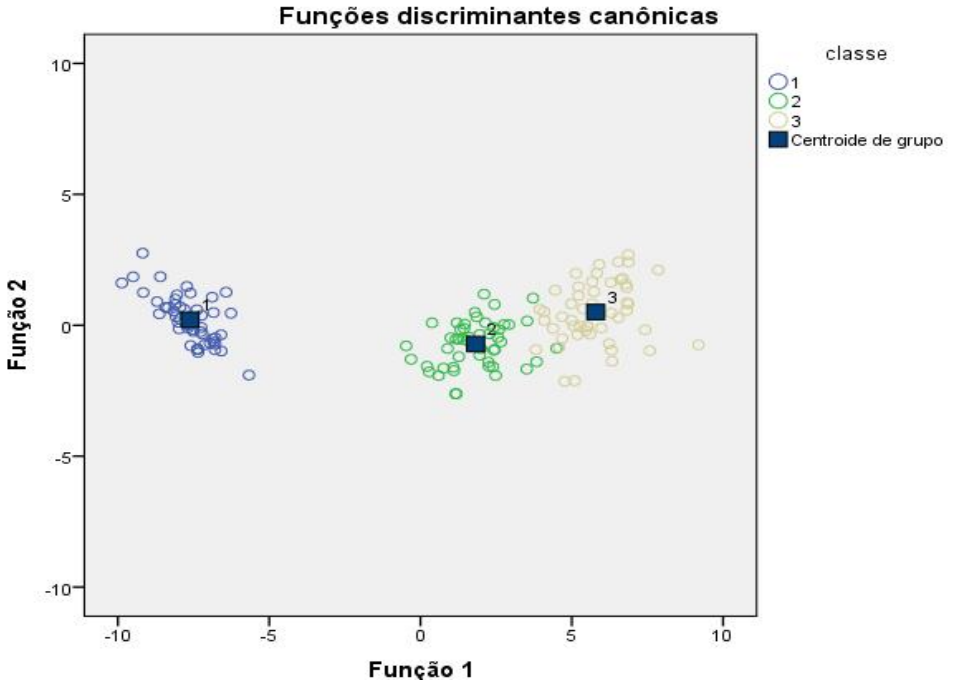


FIGURA 18 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO IRIS
FONTE: O autor (2015)

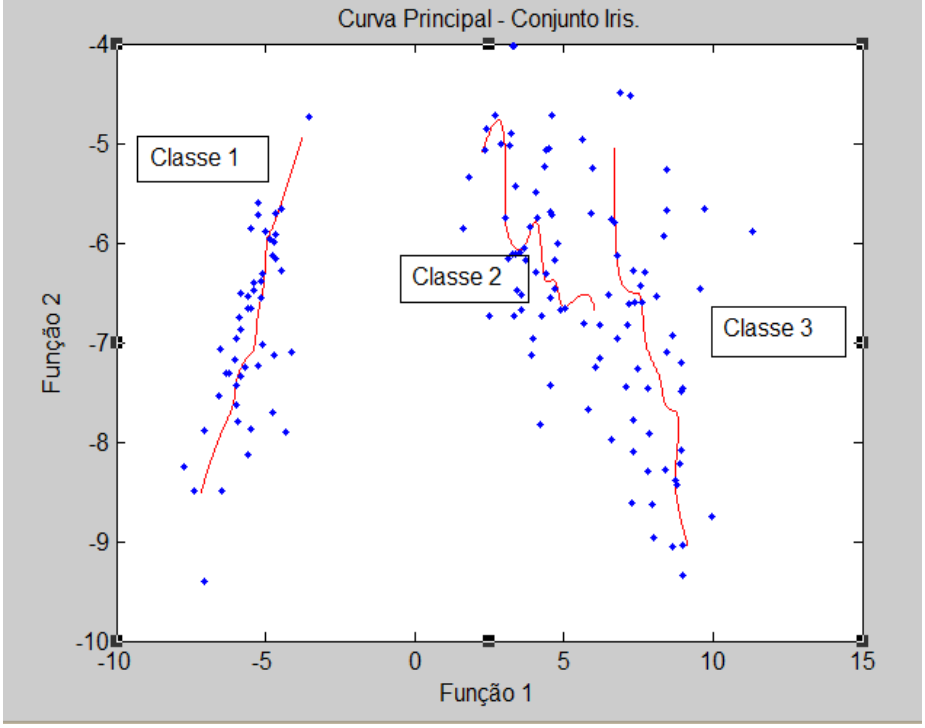


FIGURA 19 - CURVA PRINCIPAL PARA O CONJUNTO IRIS
FONTE: O autor (2015)

4.2.2 Medical

Este conjunto é formado por 54 observações, coletadas em uma pesquisa na área médica, cujo estudo refere-se a resultados clínicos de um grupo de 54 participantes que responderam a três testes. O conjunto *medical* foi obtido de Dugard, Todman e Staines (2009), Conjunto de dados para psicologia e medicina. O método de Fisher é mais eficiente apenas na classificação para a classe 2. No geral, o método *k*-segmentos é superior, com 92,5% contra 88,8% do método dos centroides, conforme resultados apresentados na tabela 5.

TABELA 4 - MATRIZ DE CONFUSÃO PARA O CONJUNTO MEDICAL PARA FDA CENTROIDES X FDA *K*-SEGMENTOS

		FDA centroides			**FDA *k*-segmentos *k=2***		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	26	**22**	3	1	**25**	1	0
		85%	12%	4%	**96%**	4%	0%
2	18	0	**17**	1	1	**15**	2
		0,00%	**94,4%**	5,56%	5,5%	**83,3%**	11,1%
3	10	0	1	**9**	0	0	**10**
		0%	10%	**90%**	0%	0%	**100%**
Casos classificados corretamente: 88,8%							**92,5%**

FONTE: O autor (2015).

Semelhante ao conjunto Iris, a transformação dos dados pela análise discriminante é bastante eficiente, conforme mostra a figura 20. Com a transformação da FDA sobre o conjunto, obtiveram-se classes efetivamente separadas e com os seus respectivos centroides no centro de seu grupo, com boa distância entre eles. Nas figuras 20 e 21 pode-se observar que devido à interseção das classes 1 e 2 e também pela maior dispersão dos pontos na classe 1 gerou maior número de classificações incorretas pela FDA. Na classe 2, o algoritmo *k*-segmentos perde eficiência devido à dispersão de forma circular. Esta forma de dispersão aumenta a distância da CP aos pontos, afetando a qualidade de classificação.

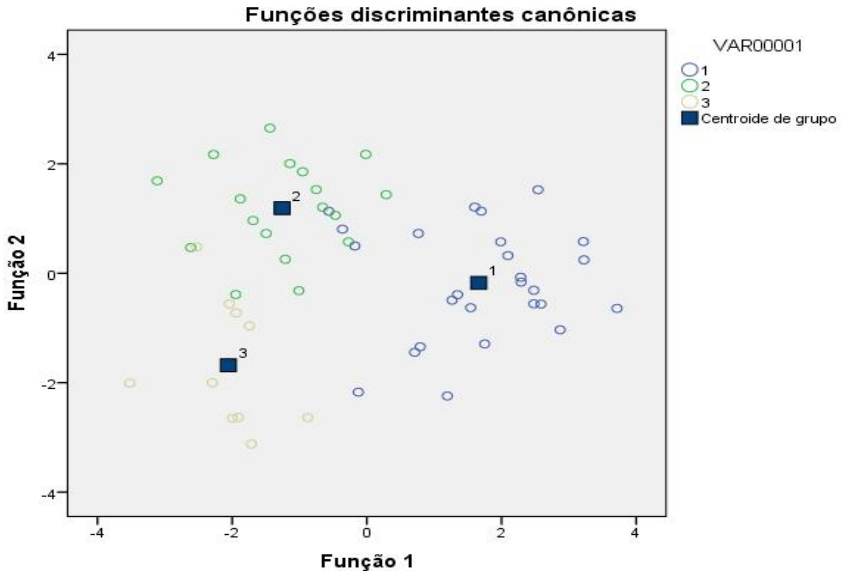


FIGURA 20 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO MEDICAL
FONTE: O autor (2015)

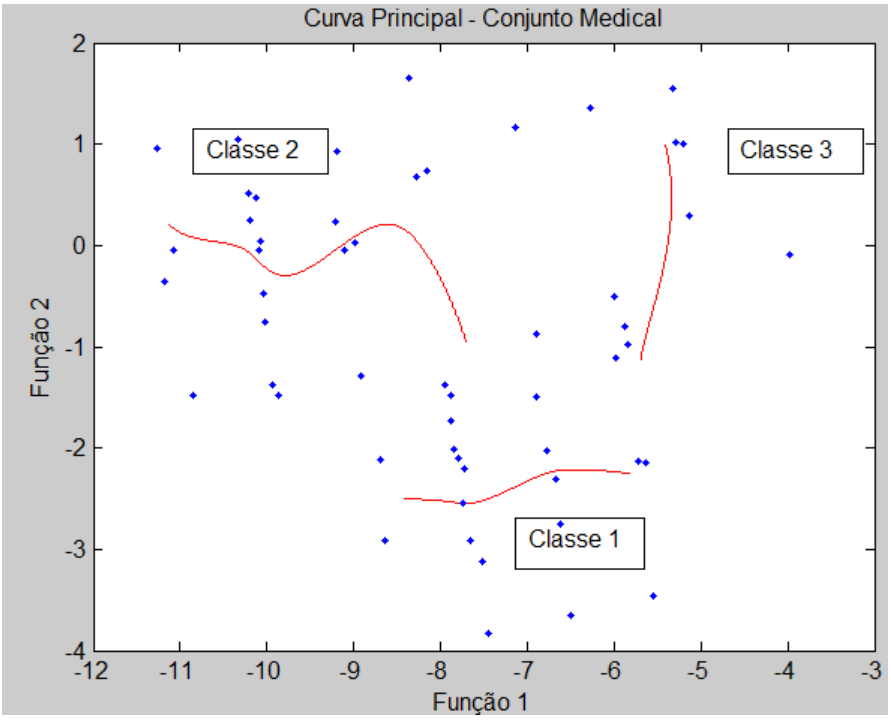


FIGURA 21 - CURVA PRINCIPAL PARA O CONJUNTO MEDICAL
FONTE: O autor (2015)

4.2.3 Wave

Este é um conjunto artificial de três classes de formas de ondas, gerado por um programa em linguagem C, obtido por Breiman *et al*. (1984). Cada classe é constituída por uma combinação convexa aleatória de duas formas de ondas, amostradas com ruído adicionado. Este conjunto contém 5000 observações e 21 variáveis.

Na análise da matriz de confusão, tabela 6, é visto que o método *k*-segmentos obteve resultado pior que o método dos centroides. Ao analisar individualmente o desempenho por classe pode-se observar que o algoritmo *k*-segmentos foi inferior na classificação nas classes 2 e 3. A figura 22 mostra a CP para este conjunto, com interseções entre as classes, os escorres estão concentrados. Este aglomerado de pontos prejudica a eficiência do algoritmo *k*-segmentos.

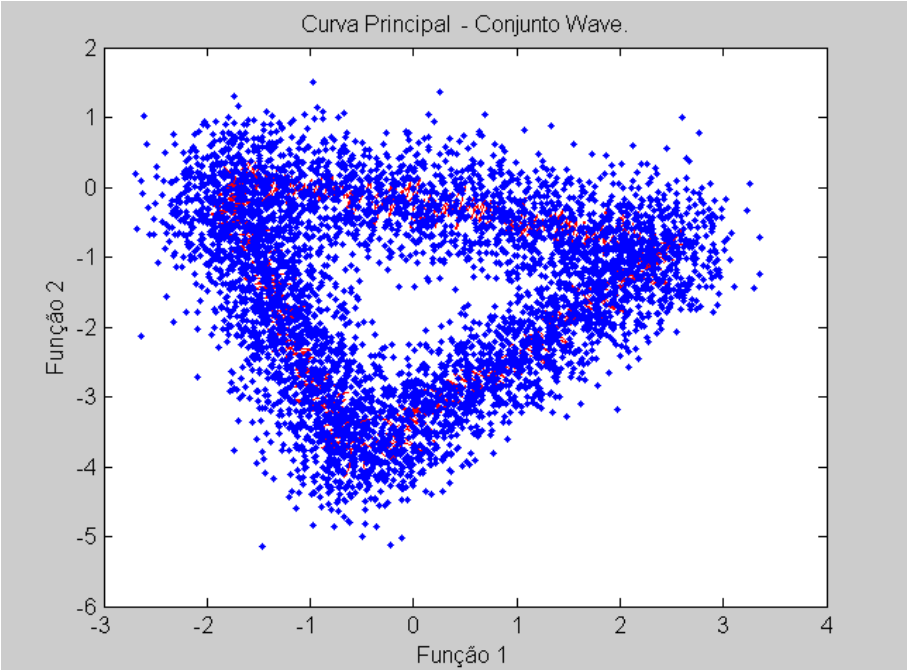


FIGURA 22 - CURVA PRINCIPAL PARA O CONJUNTO WAVE
FONTE: O autor (2015)

TABELA 5 - MATRIZ DE CONFUSÃO PARA O CONJUNTO *WAVE* PARA FDA CENTROIDES X FDA *K*-SEGMENTOS

FDA centroides					**FDA *k*-segmentos *k=5***		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	1657	**1327**	152	178	**1388**	134	135
		80,08%	9,17%	10,74%	**83,77%**	8,09%	8,15%
2	1647	103	**1451**	93	151	**1423**	73
		6,25%	**88,10%**	5,65%	9,17%	**86,40%**	4,43%
3	1696	75	71	**1550**	179	113	**1404**
		4,4%	4,2%	**91,4%**	10,55%	6,66%	**82,78%**
Casos classificados corretamente: 86,56%					**84,30%**		

FONTE: O autor (2015)

A figura 23 mostra as classes com grande interseção nos extremos das classes. De modo geral, a discriminação dos conjuntos foi ótima com 86,56% de eficiência na classificação dos dados (FDA por centroides).

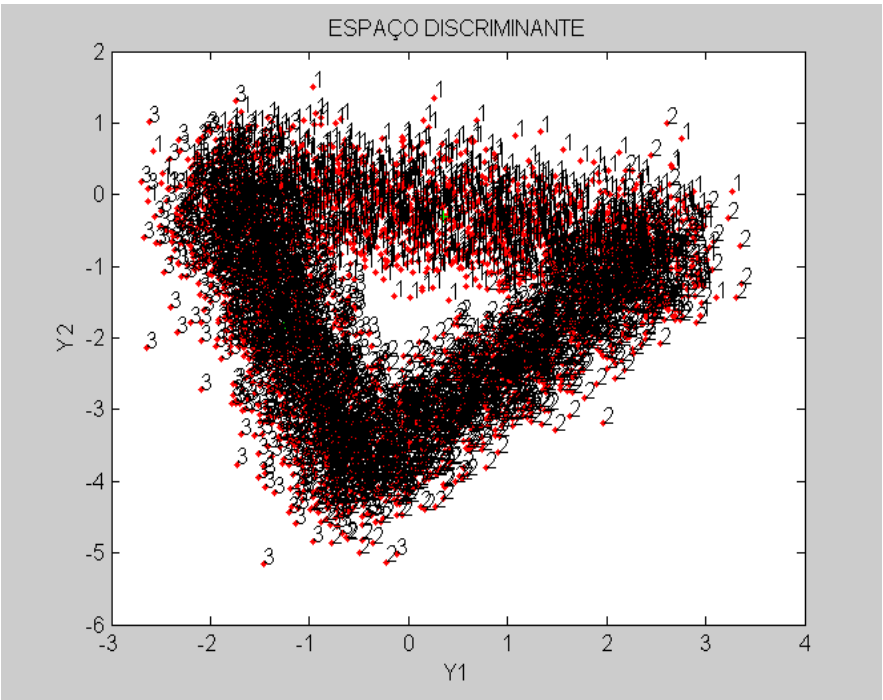


FIGURA 23 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO WAVE
FONTE: O autor (2015)

4.2.4 Wine

O conjunto de dados wine contém 3 classes de dados. A classe 1 tem 59 observações amostrais, a classe 2 tem 71 e a classe 3 tem 48 conjuntos de dados. O número de variáveis é 13: álcool, MalicAcid*, Ash,* AlcalinityOfAsh*,* magnésio*,* TotalPhenols*,* Flavanoids*,* NonflavanoidPhenols*,* Proanthocyanins*,* COLORIN-*intensidade, matiz,* OD280/OD315 e prolina*.* O algoritmo *k*-segmentos e o método de Fisher apresentam 100% de classificação correta dos dados, conforme apresentado no tabela 7.

TABELA 6 - MATRIZ DE CONFUSÃO PARA O CONJUNTO *WINE* PARA FDA CENTROIDES X FDA *K*-SEGMENTOS

FDA centroides					**FDA *k*-segmentos *k=2***		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	59	**59**	0	0	**59**	0	0
		100%	0%	0%	**100%**	0%	0%
2	71	0	**71**	0	0	**71**	0
		0%	**100,0%**	0,00%	0%	**100,0%**	0,00%
3	48	0	0	**48**	0	0	**48**
		0%	0%	**100%**	0%	0%	**100%**
Casos classificados corretamente: 100%							**100%**

FONTE: O autor (2015)

Na análise da transformação efetuada pela análise discriminante, observa-se as classes bem separadas, sem interseção dos escores e com os centroides nos centros de suas respectivas classes, o que é desejável nas duas técnicas. Da mesma forma que ocorre com os centroides, as curvas principais passam pelo meio das classes, como pode ser observado nas figuras 24 e 25. Este fato explica a eficiência de 100% na classificação do conjunto.

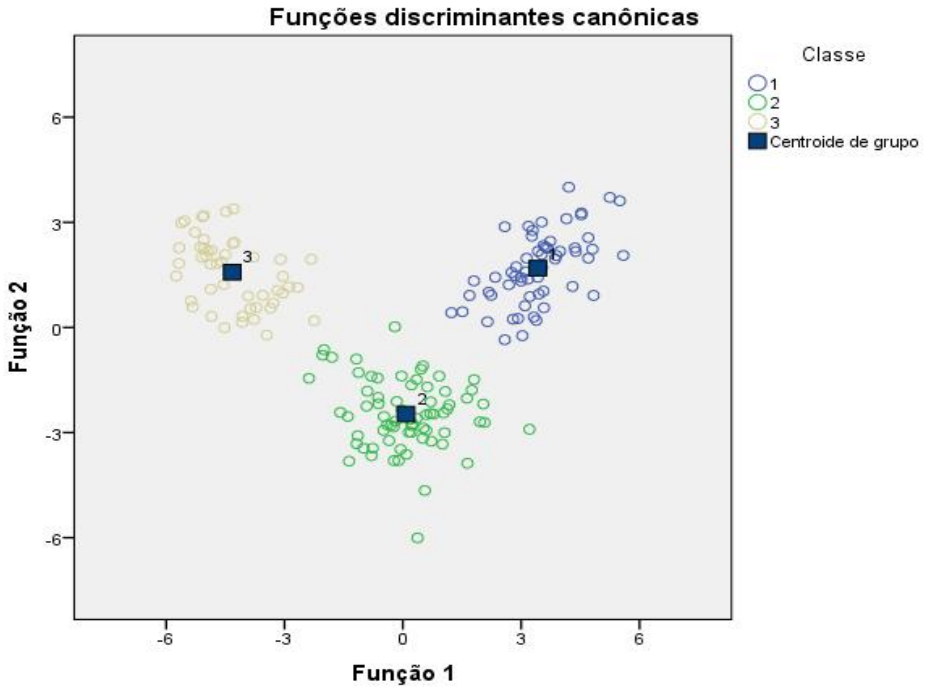


FIGURA 24 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO WINE.
FONTE: O autor (2015).

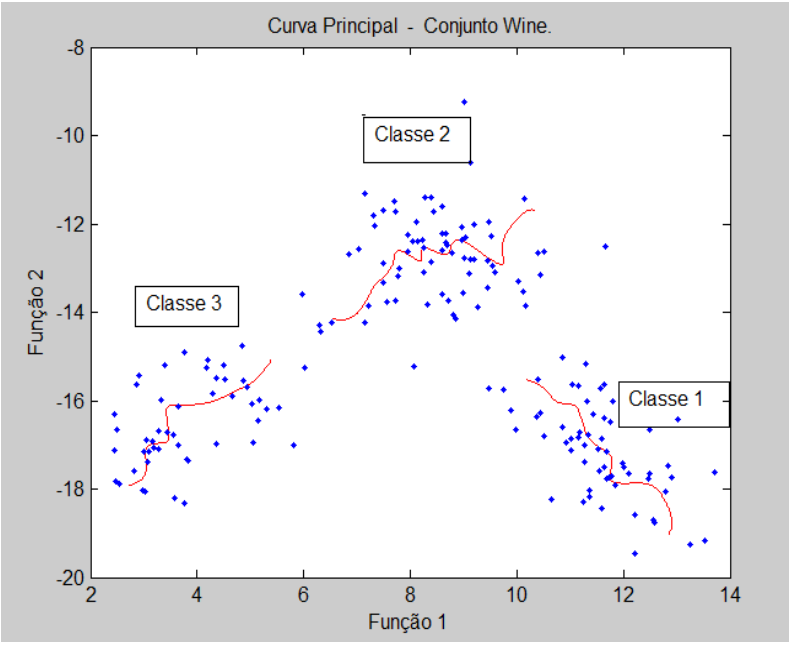


FIGURA 25 - CURVA PRINCIPAL PARA O CONJUNTO WINE
FONTE: O autor (2015)

4.2.5 Álcool

Com o arquivo de dados álcool (TANAGRA DATASETS, 2014), o objetivo é prever o tipo de álcool (*kirsch, mirab, poire*). A amostra contém 77 observações, com 6 variáveis (butanol, etc.). Na análise da matriz de confusão, tabela 8, a classe *kirsch* apresenta resultado ótimo nos dois métodos e o bom comportamento para o conjunto baseia-se principalmente sobre esta classe. As classes *Mirab* e *Poire* não obtiveram o ótimo resultado da primeira classe com classificação correta inferior a 80% pelo método dos centroides. Já o método *k*-segmentos foi superior, tanto individualmente por classes, como no resultado geral com no mínimo 86,2% de acerto.

TABELA 7 - MATRIZ DE CONFUSÃO PARA O CONJUNTO ÁLCOOL PARA FDA CENTROIDES X FDA *K*-SEGMENTOS

FDA centroides					**FDA *k*-segmentos *k=4***		
Classe	**Tamanho da Classe**	**Classe predita**			**Classe predita**		
		Kirsch	***Mirab***	***Poire***	***Kirsch***	***Mirab***	***Poire***
Kirsch	18	**18**	0	0	**18**	0	0
		100%	0%	0%	**100,00%**	0,00%	0,00%
Mirab	29	0	**23**	6	0	**25**	4
		0%	**79,31%**	20,69%	0,00%	**86,21%**	13,79%
Poire	30	0	9	**21**	0	4	**26**
		0%	30%	**70%**	0,00%	13,33%	**86,67%**
Casos classificados corretamente: 80,52%					**89,61%**		

FONTE: O autor (2015)

Na análise das figuras 26 e 27, gráficos dos centroides e da curva principal, é fácil ver que o desempenho na classificação é inferior nas classes *Mirab* e *Poire*, pois os escores da classe *Kirsch* estão separados das classes *Mirab* e *Poire*, porém a interseção dos escores destas classes são grandes, pois o espaço discriminante está sobreposto. A classificação foi menos eficiente nas classes *Mirab* e *Poire*, também, devido à proximidade dos centroides (FDA) das curvas principais (*k*-segmentos) e ao cruzamento das curvas principais (para o algoritmo *k*-segmentos).

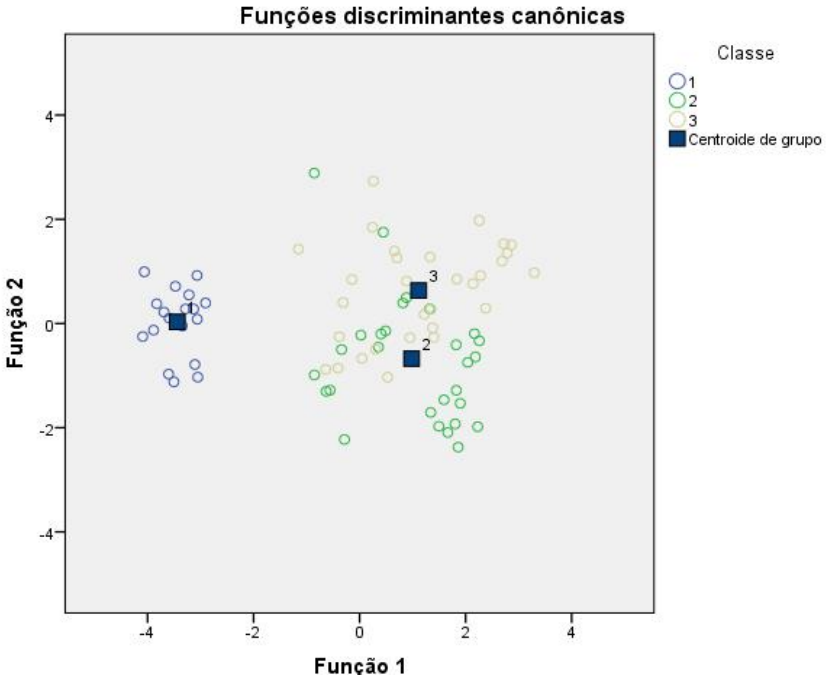


FIGURA 26 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO ÁLCOOL
FONTE: O autor (2015)

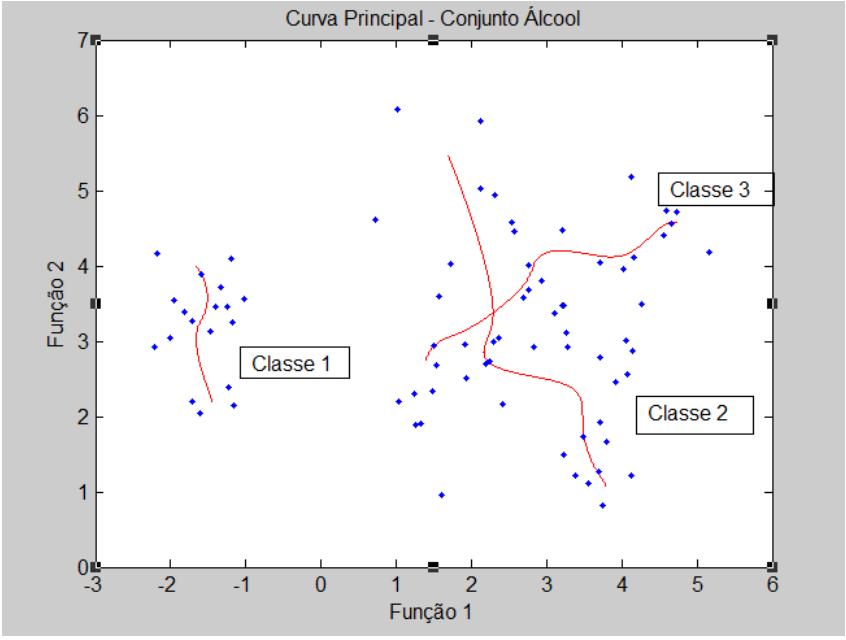


FIGURA 27 - CURVA PRINCIPAL PARA O CONJUNTO ÁLCOOL
FONTE: O autor (2015)

4.2.6 Tiroide

Neste conjunto composto por 3 classes, o problema é determinar se um paciente tem a tiroide no estado normal ou não (hipertiroidismo ou hipotiroidismo). O diagnóstico (o rótulo de classe) foi baseado em um informe médico completo, incluindo *anamnese*, exame, etc. O bom desempenho dos dois métodos na classificação global (94,42% e 97,21%) pode ser observado na tabela 9, principalmente pelos resultados obtidos na primeira e terceira classe com 99% e 100%, respectivamente, de classificação correta nas duas técnicas. A principal diferença de desempenho ocorre na 2ª classe, com vantagem do algoritmo *k*-segmentos.

TABELA 8 - MATRIZ DE CONFUSÃO PARA O CONJUNTO TIROIDE PARA FDA CENTROIDES X FDA K-SEGMENTOS

		FDA centroides			**FDA *k*-segmentos *k=4***		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	150	**149**	1	0	**147**	2	1
		99%	1%	0%	**98%**	1%	1%
2	35	5	**30**	0	3	**32**	0
		14,29%	**85,71%**	0,00%	9%	**91%**	0%
3	30	6	0	**24**	0	0	**30**
		20%	0%	**100%**	0%	0%	**100%**
Casos classificados corretamente: 94,42%					**97,21%**		

FONTE: O autor (2015)

As figuras 28 e 29, mostram a concentração dos escores da classe 1 em torno do centroide, o que pode afetar o desempenho do algoritmo *k*-segmentos. O resultado é que o algoritmo não é superior nessa classe (99% contra 98%), enquanto para o método dos centroides esta é a concentração ideal para a sua eficiência, pois os pontos estão aglomerados em torno do centroide. Já nas classes 2 e 3, os escores estão dispersos de forma longitudinal, o que favorece o algoritmo *k*-segmentos. Para o método dos centroides, esta forma de dispersão não é a ideal, pelo fato de a mesma ter apenas um ponto por classe.

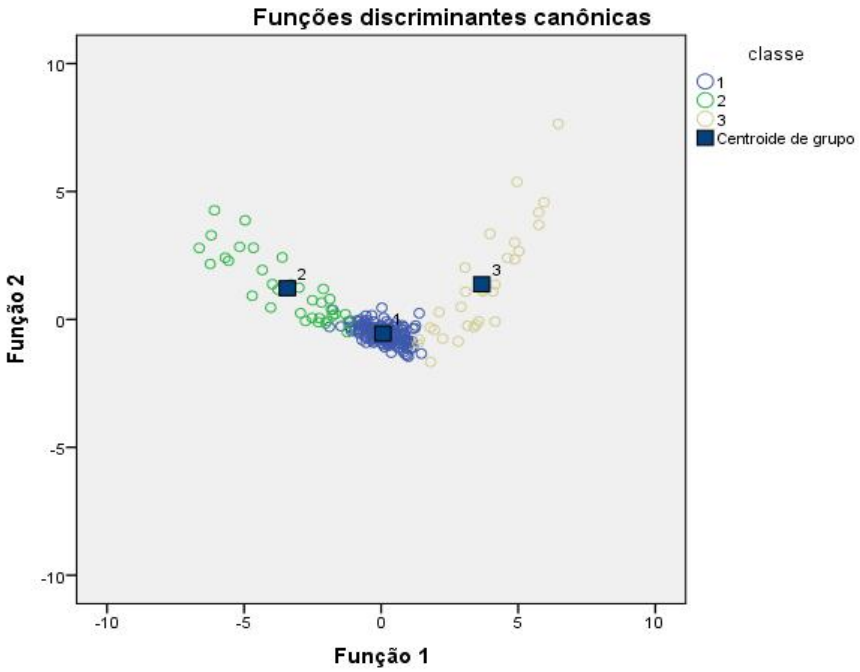


FIGURA 28 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO TIROIDE
FONTE: O autor (2015)

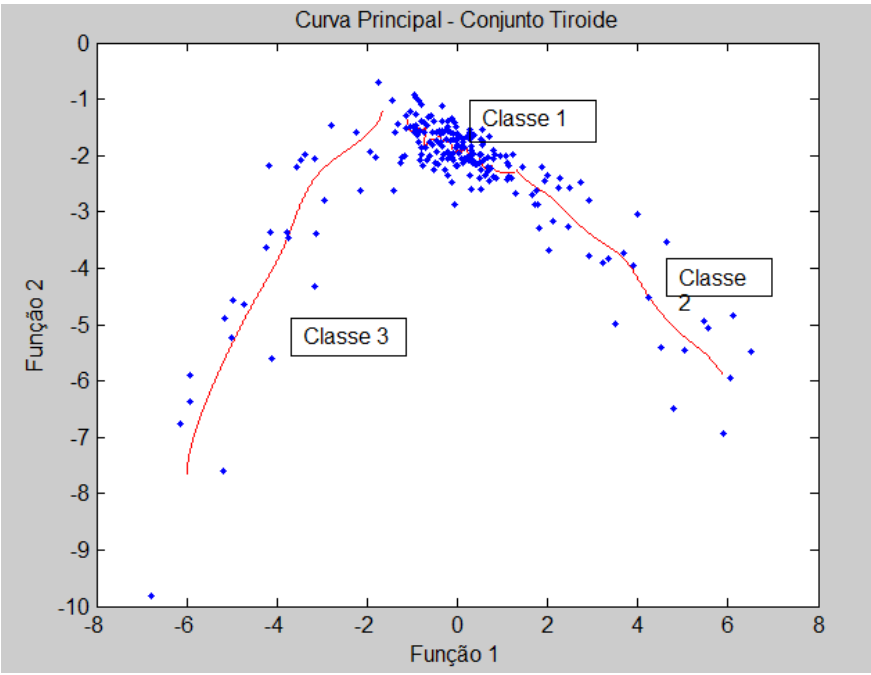


FIGURA 29 - CURVA PRINCIPAL PARA O CONJUNTO TIROIDE
FONTE: O autor (2015)

4.2.7 Glass

O conjunto *Glass* é composto por 214 amostras com 6 classes[3] e o estudo da classificação dos tipos de vidro foi motivado por investigação criminológica, sendo que o tipo de vidro, coletado na cena do crime, pode ser usado como prova. O método *k*-segmentos foi superior na classificação e nos dois métodos a classificação foi razoável (acima de 60%, para os dois métodos). A análise das percentagens de classificações incorretas, apresentadas na tabela 10, mostra que o modelo *k*-segmentos tem melhor capacidade preditiva, embora o desempenho individual na classe 2 seja inferior ao obtido pelos centroides.

TABELA 9 - MATRIZ DE CONFUSÃO PARA O CONJUNTO GLASS PARA FDA CENTROIDE X FDA K-SEGMENTOS

FDA centroides						**FDA *k*-segmentos *k=5***					
Classe predita						**Classe predita**					
1	**2**	**3**	**5**	**6**	**7**	**1**	**2**	**3**	**5**	**6**	**7**
46	14	10	0	0	0	**52**	10	8	0	0	0
65,7%	20,0%	14,3%	0,0%	0,0%	0,0%	**74,29%**	14,29%	11,43%	0,00%	0,00%	0,00%
16	**41**	12	4	3	0	24	**40**	7	2	2	1
21,1%	**54,0%**	15,8%	5,3%	4,0%	0,0%	31,58%	**52,63%**	9,21%	2,63%	2,63%	1,32%
3	3	**11**	0	0	0	0	1	**16**	0	0	0
17,7%	17,7%	**64,7%**	0,0%	0,0%	0,0%	0,00%	5,88%	**94,12%**	0,00%	0,00%	0,00%
0	2	0	**10**	0	1	0	0	0	**13**	0	0
0,0%	15,4%	0,0%	**76,9%**	0,0%	7,7%	0,00%	0,00%	0,00%	**100,00%**	0,00%	0,00%
1	1	0	0	**7**	0	0	1	0	1	**7**	0
11,1%	11,1%	0,0%	0,0%	**77,8%**	0,0%	0,00%	11,11%	0,00%	11,11%	**77,78%**	0,00%
0	1	1	2	1	**24**	1	0	0	0	1	**27**
0,0%	3,5%	3,5%	6,9%	3,5%	**82,8%**	3,45%	0,00%	0,00%	0,00%	3,45%	**93,10%**
Casos classificados corretamente: 64,95%						**72,42%**					

FONTE: O autor (2015)

Na análise das figuras 30 e 31, verifica-se que há grande interseção entre os escores de cada classe, principalmente nas classes 1, 2 e 3. Na classificação por centroides estas três classes obtiveram o pior desempenho. Já o método *k*-segmentos, também afetado pela proximidade dos escores dessas classes, teve

[3] O conjunto *Glass* contêm 7 classes, porém a classe 4 não consta nenhum elemento observado.

desempenho ruim na classe 2 (52,63%), relacionando elementos desse grupo nas classes 1 e 3, resultado que afetou a classificação geral por este método. Nas demais classes, o método *k*-segmentos teve desempenho muito superior.

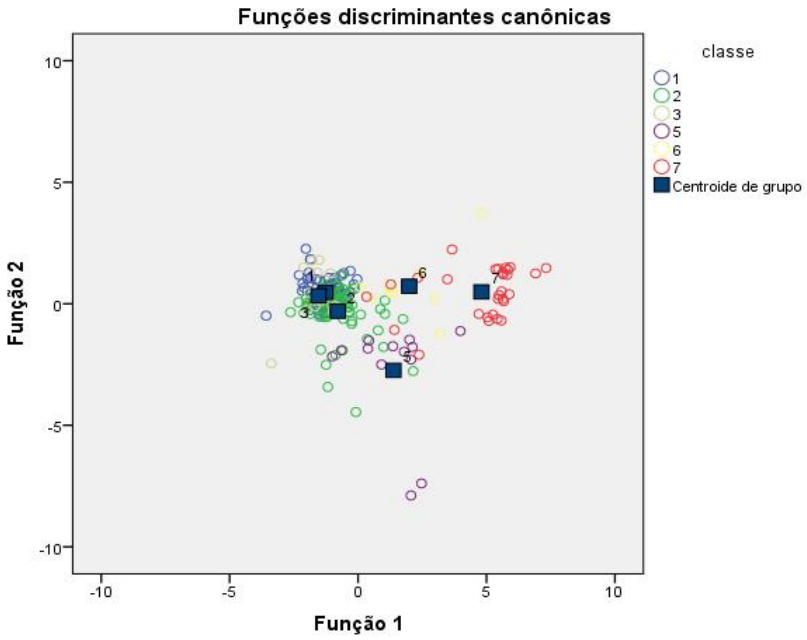


FIGURA 30 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO GLASS
FONTE: O autor (2015)

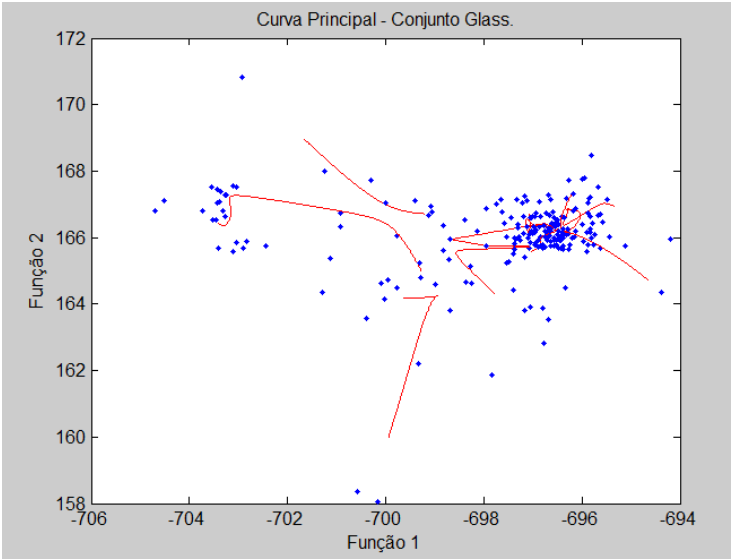


FIGURA 31 - CURVA PRINCIPAL PARA O CONJUNTO GLASS
FONTE: O autor (2015)

4.2.8 Futebol

Os dados deste conjunto foram coletados por G. R. Bryce e R. M. Barker (*Brigham Young University*) e estão disponíveis na UCI *Machine Learning*. Este conjunto faz parte de um estudo preliminar de uma possível ligação entre desenho do capacete de futebol e lesões no pescoço. Seis medidas da cabeça foram feitas em cada indivíduo. Havia 30 indivíduos em cada um dos três grupos: jogadores de futebol da escola (grupo 1), jogadores de futebol da faculdade (grupo 2), e os jogadores que não estudam (grupo 3). As seis variáveis são: largura da cabeça, circunferência da cabeça, medição ao nível dos olhos para frente e para trás, medição do olho ao topo da cabeça, orelha ao topo da cabeça e da largura da mandíbula. A matriz de confusão, tabela 11, mostra os resultados obtidos com este conjunto. A eficiência foi razoável, com 73,33% por centroides e 76,67% para o algoritmo *k*-segmentos, na classificação correta dos vetores.

TABELA 10 - MATRIZ DE CONFUSÃO PARA O CONJUNTO FUTEBOL PARA FDA CENTROIDES X FDA *K*-SEGMENTOS

FDA centroides					**FDA *k*-segmentos *k=4***		
Classe	**Classe predita**				**Classe predita**		
	Tamanho da classe	**1**	**2**	**3**	**1**	**2**	**3**
1	30	**26**	1	3	**26**	0	4
		86,67%	3,33%	10,00%	**86,67%**	0,00%	13,33%
2	30	1	**20**	9	3	**20**	7
		3,33%	**66,67%**	30,00%	10,00%	**66,67%**	23,33%
3	30	2	8	**20**	3	4	**23**
		6,67%	26,67%	**66,67%**	10,00%	13,33%	**76,67%**
% de acertos:		**73,33%**			% de acertos:		**76,67%**

FONTE: O autor (2015)

Devido à interseção entre os escores, principalmente das classes 2 e 3 e também devido à proximidade dos centros, o método obteve um resultado razoável na classificação geral. Nas classes 2 e 3, obteve eficiência abaixo da média geral, quando utilizado o método dos centroides. A classe 1, melhor discriminada das

demais, obteve melhor eficiência (86,67%), conforme é apresentado na tabela 11 e na figura 32.

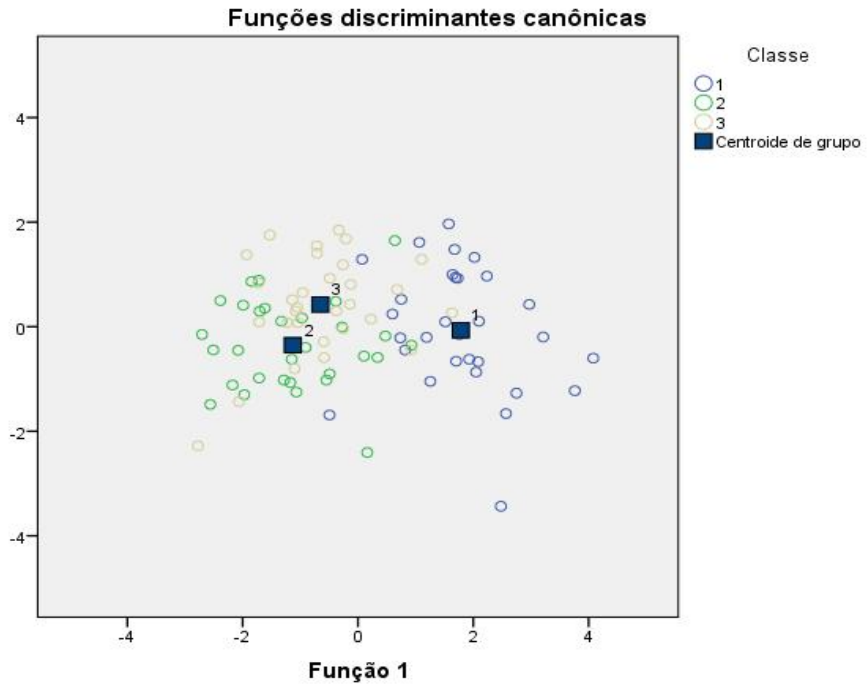


FIGURA 32 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO FUTEBOL
FONTE: O autor (2015)

Os dois métodos apresentaram resultado muito semelhantes. Nas classes 1 e 2, os resultados foram iguais e apenas na classe 3 o método *k*-segmento foi superior. O método *k*-segmentos perdeu um pouco de sua eficiência, devido a interseção das curvas principais nas classes 2 e 3 (FIGURA 33).

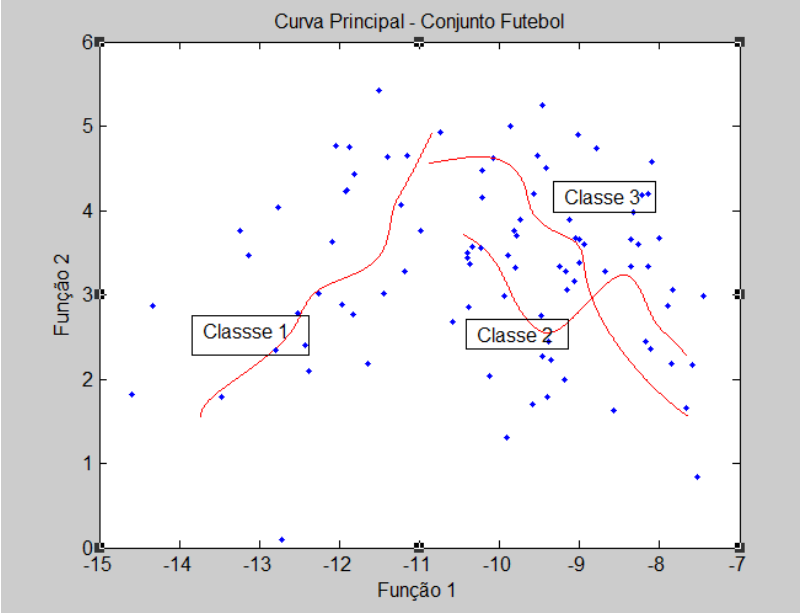


FIGURA 33 - CURVA PRINCIPAL PARA O CONJUNTO FUTEBOL
FONTE: O autor (2015)

4.2.9 Segment

Este conjunto é referente à segmentação de imagens obtidas de um banco de dados, criado pela *Vision Group*, da Universidade de Massachusetts. O conjunto é formado por 19 variáveis – sendo que as variáveis 4, 5 e 6 foram retiradas do conjunto pelo fato de gerar matrizes de covariância singulares, (as variáveis removidas apresentam grande quantidade de zeros) – 7 classes (tipos de imagens) e 2310 observações amostrais. Na análise da matriz de confusão (TABELAS 12 e 13), os dois métodos apresentaram resultados semelhantes, com ótimo desempenho (acima de 90% de acerto). As classes 6 e 7 apresentaram desempenho abaixo da média geral (principalmente a classe 7). O Resultado individual também apresentou resultado muito semelhante nos dois métodos.

TABELA 11 - MATRIZ DE CONFUSÃO PARA O CONJUNTO SEGMENT PARA FDA CENTROIDE

		FDA centroides						
Classe	**Tamanho**	**Classe predita**						
	da classe	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	330	**330**	0	0	0	0	0	0
		100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2	330	0	**295**	0	0	2	3	30
		0,00%	**89,39%**	0,00%	0,00%	0,61%	0,91%	9,09%
3	330	0	0	**330**	0	0	0	0
		0,00%	0,00%	**100,00%**	0,00%	0,00%	0,00%	0,00%
4	330	2	0	0	**327**	0	0	1
		0,61%	0,00%	0,00%	**99,09%**	0,00%	0,00%	0,30%
5	330	0	0	0	0	**324**	0	6
		0,00%	0,00%	0,00%	0,00%	**98,18%**	0,00%	1,82%
6	330	19	0	0	0	1	**273**	37
		5,76%	0,00%	0,00%	0,00%	0,30%	**82,73%**	11,21%
7	330	0	70	0	0	2	17	**241**
		0,00%	21,21%	0,00%	0,00%	0,61%	5,15%	**73,03%**
		% de acertos:		**91,77%**				

FONTE: O autor (2015)

TABELA 12 - MATRIZ DE CONFUSÃO PARA O CONJUNTO SEGMENT PARA FDA K-SEGMENTOS

		FDA *k*-segmentos *k=4*						
Classe	**Tamanho da classe**	**Classe predita**						
		1	**2**	**3**	**4**	**5**	**6**	**7**
1	330	**330**	0	0	0	0	0	0
		100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2	330	0	**291**	0	0	3	4	32
		0,00%	**88,18%**	0,00%	0,00%	0,91%	1,21%	9,70%
3	330	0	0	**330**	0	0	0	0
		0,00%	0,00%	**100,00%**	0,00%	0,00%	0,00%	0,00%
4	330	0	0	0	**330**	0	0	0
		0,00%	0,00%	0,00%	**100,00%**	0,00%	0,00%	0,00%
5	330	0	0	0	0	**330**	0	0
		0,00%	0,00%	0,00%	0,00%	**100,00%**	0,00%	0,00%
6	330	24	0	0	0	16	**286**	4
		7,27%	0,00%	0,00%	0,00%	4,85%	**86,67%**	1,21%
7	330	0	46	0	0	17	26	**241**
		0,00%	13,94%	0,00%	0,00%	5,15%	7,88%	**73,03%**
		% de acertos:		**92,55%**				

FONTE: O autor (2015)

As figuras 34 e 35 mostram as classes 1, 2, 5, 6 e 7 aglomeradas, com os centroides próximos, o que não ocorre com as classes 3 e 4. Esta aglomeração afetou o resultado final apenas das classes 2, 6 e 7, o que prejudicou a eficiência na classificação geral.

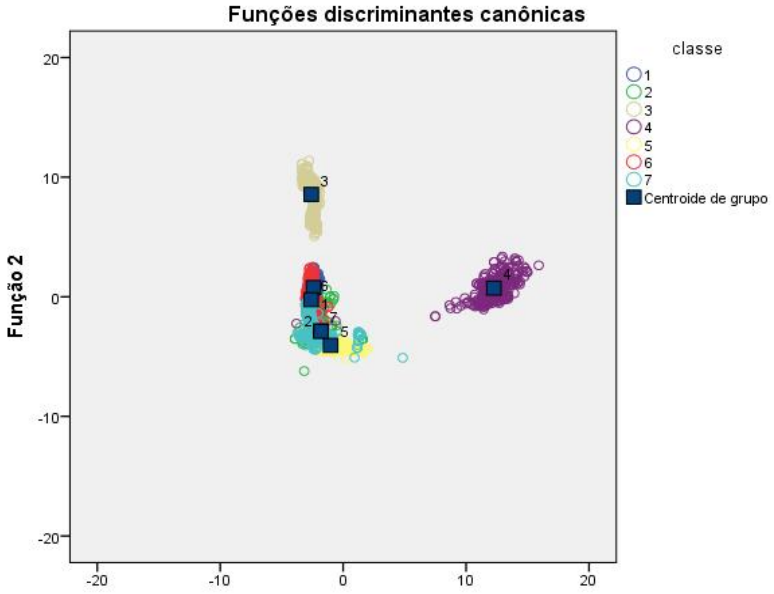


FIGURA 34 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO SEGMENT
FONTE: O autor (2015)

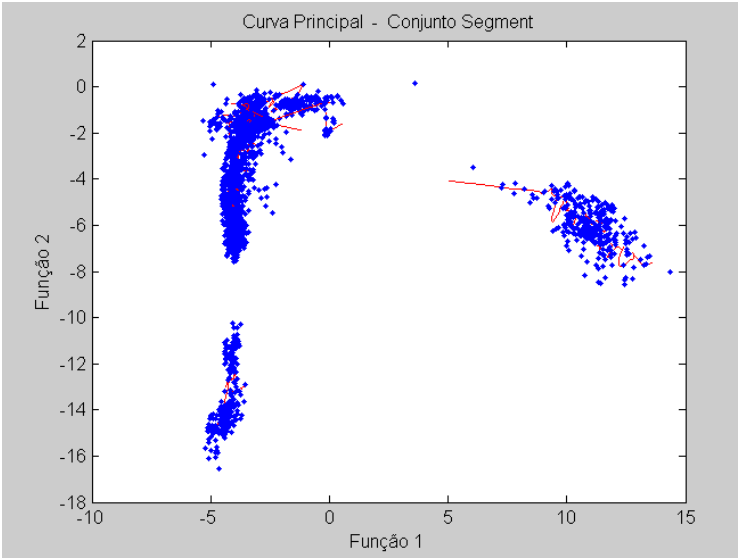


FIGURA 35 - CURVA PRINCIPAL PARA O CONJUNTO SEGMENT
FONTE: O autor (2015)

4.2.10 Besouro

Os dados deste conjunto contêm informações físicas de 74 besouros de três espécies de besouro saltador (*Ch. Concinna, Ch. Heptapotamica, e Ch. Heikertingeri*), que formam as classes desse conjunto, com 6 variáveis. Na análise da matriz de confusão (TABELA 14), é fácil ver que a discriminação das 3 classes foi totalmente eficiente. As classes ficaram separadas uma das outras sem nenhuma interseção, o que foi ótimo para comparação das distâncias dos centroides ao vetor aleatório candidato a classificação.

TABELA 13- MATRIZ DE CONFUSÃO PARA O CONJUNTO BESOURO PARA FDA CENTROIDE X FDA K-SEGMENTOS

FDA centroides					**FDA *k*-segmentos *k=2***		
Classe	**Classe predita**				**Classe predita**		
	Tamanho da classe	**1**	**2**	**3**	**1**	**2**	**3**
1	21	**21**	0	0	**21**	0	0
		100%	0%	0%	**100,00%**	0,00%	0,00%
2	22	0	**22**	0	0	**22**	0
		0%	**100%**	0%	0,00%	**100,00%**	0,00%
3	31	0	0	**31**	0	0	**31**
		0%	0%	**100%**	0,00%	0,00%	**100,00%**
Casos classificados corretamente: 100%					**Casos classificados corretamente: 100%**		

FONTE: O autor (2015)

As duas técnicas obtiveram 100% de eficiência, como mostram as figuras 36 e 37. Do mesmo modo como ocorre com os centroides, a curva principal de cada classe 'passa' pelo centro da classe, o que é favorável para o método *k*-segmentos, como também é favorável a não ocorrência de interseção entre as classes.

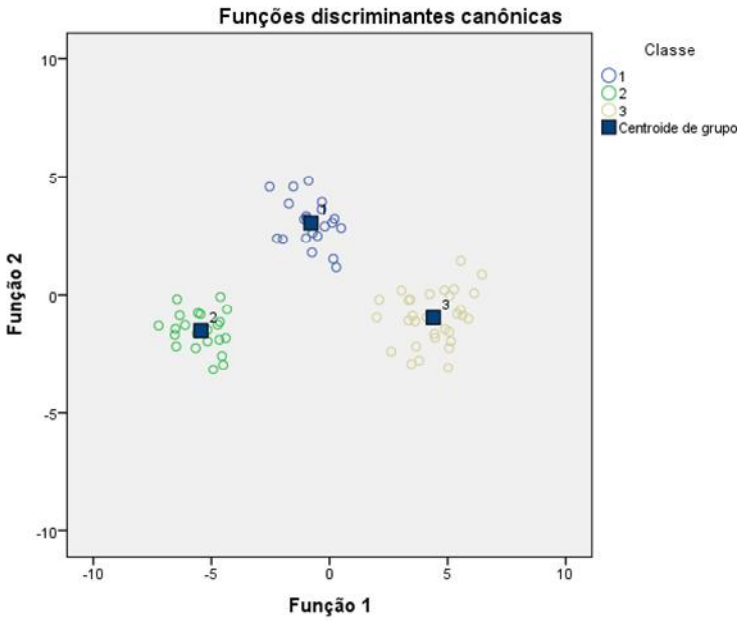


FIGURA 36 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO BESOURO
FONTE: O autor (2015)

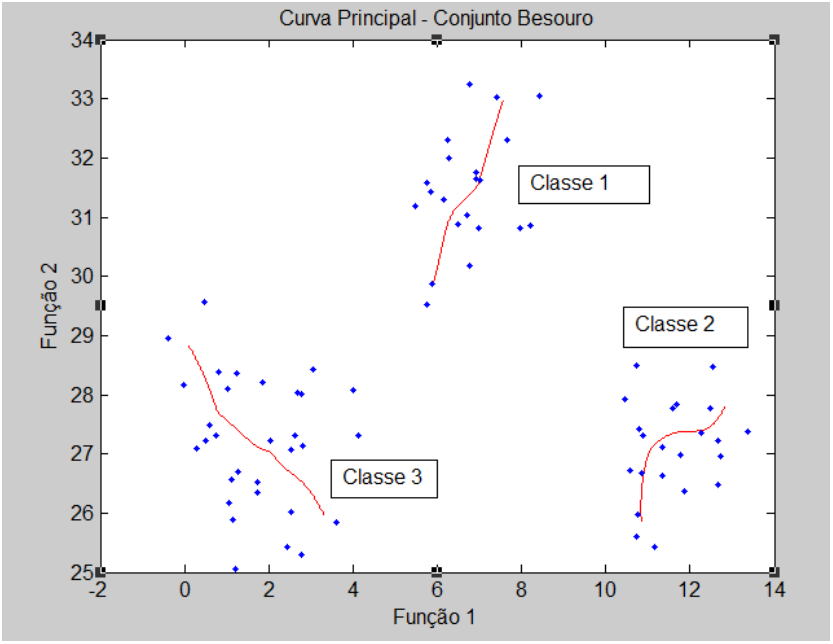


FIGURA 37 - CURVA PRINCIPAL PARA O CONJUNTO BESOURO
FONTE: O autor (2015)

4.2.11 Abalone

Abalone (*haliotis*) é um molusco que vive em concha e possui diversas espécies, sua carne é muito apreciada na Ásia. Cada observação é composta por 8 variáveis: comprimento (mm), diâmetro (mm), altura (mm), peso inteiro (g), peso da carne (g), peso das vísceras (g), peso da concha (g) e anéis (anos). As classes são: Macho (M), Fêmea (F) e Imaturo (I). O objetivo é classificar uma nova observação quanto ao sexo do indivíduo.

A separação das classes pela análise discriminante não foi muito eficiente para as classes 1 e 2. A análise dos centroides das classes explica este resultado, pois nas classes 1 e 2 os centroides estão muito próximos e dentro da classe 3 (FIGURA 38), adiante. Analisando a matriz de confusão (TABELA 15), o método dos centroides classificou incorretamente 43,19% das observações da classe 1 na classe 2 e com 22,64% de classificações incorretas na terceira classe. Somente na classe 3 que a identificação foi mais eficiente (78,84%), pois o centroide da 3ª classe está no centro de sua classe. O método *k*-segmentos apresentou resultado semelhante ao dos centroides somente na classe 3, com eficiência de 78,84% e 70,04% respectivamente. Este resultado é inferior ao dos centroides, com péssima eficiência.

TABELA 14 - MATRIZ DE CONFUSÃO PARA O CONJUNTO ABALONE PARA FDA CENTROIDE X FDA *K*-SEGMENTOS

FDA centroides					**FDA *k*-segmentos *k=2***		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	1528	**522**	660	346	**681**	367	480
		34,16%	43,19%	22,64%	**44,57%**	24,02%	31,41%
2	1307	414	**689**	204	567	**383**	357
		31,68%	**52,72%**	15,61%	43,38%	**29,30%**	27,31%
3	1342	154	130	**1058**	311	91	**940**
		11,48%	9,69%	**78,84%**	23,17%	6,78%	**70,04%**
Casos classificados corretamente: 54,32%					**Casos classificados corretamente: 47,97%**		

FONTE: O autor (2015)

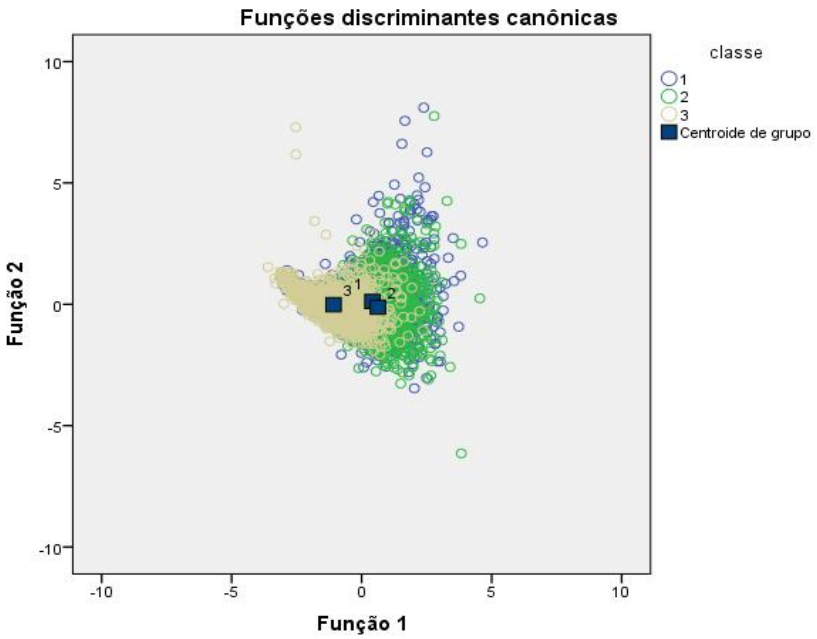


FIGURA 38 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO ABALONE
FONTE: O autor (2015)

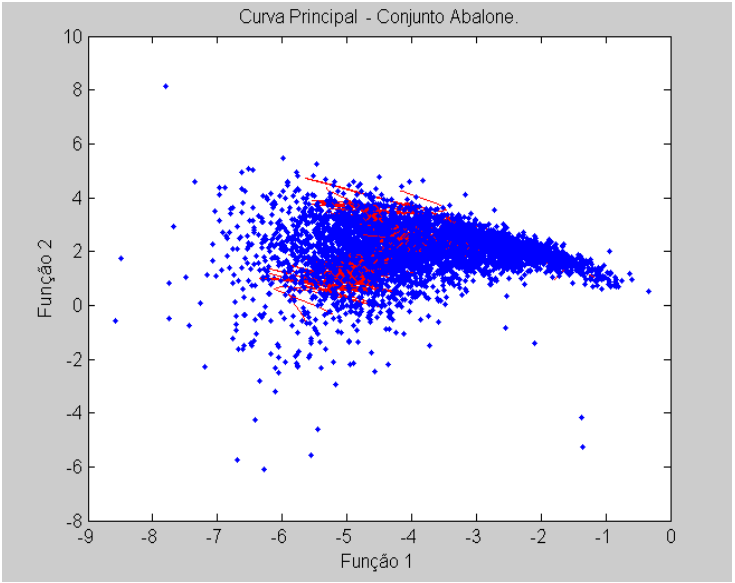


FIGURA 39 - CURVA PRINCIPAL PARA O CONJUNTO ABALONE
FONTE: O autor (2015)

Ao analisar a linha poligonal gerada pelo algoritmo, este não foi eficiente no ajuste da linha, devido às classes não estarem bem discriminadas. Observa-se na figura 39 (página anterior) que o algoritmo gerou segmentos cruzados, devido à aglomeração dos escores. Como o conjunto dos escores discriminantes está compactado, isso prejudica o ajuste de curva principal e, como consequência, o algoritmo *k*-segmentos tem desempenho inferior ao dos centroides na classificação geral e individualmente por classes.

4.2.12 Letter

O Conjunto foi desenvolvido por Frey e Slate (1991) que investigou a capacidade da técnica holland de classificar corretamente letras com 20 variações diferentes quanto a forma. O conjunto é formado por 20.000 observações e 26 classes, no qual cada classe representa uma letra do nosso alfabeto. Por exemplo, a classe 1 a letra a, classe 2 a letra b, …, a classe 26 a letra z. A forma de cada letra foi distorcida aleatoriamente em 20 fontes diferentes. Na análise gráfica dos centroides (FIGURA 40) os conjuntos de escores de cada classe estão aglomerados, com algumas classes com forte interseção de seus escores e com centroides próximos. Por estes motivos o resultado na classificação pelo método dos centroides obteve apenas 70,4 % de eficiência e, em particular em algumas classes, o desempenho foi inferior a 50% (apêndice 1, 2 e 3).

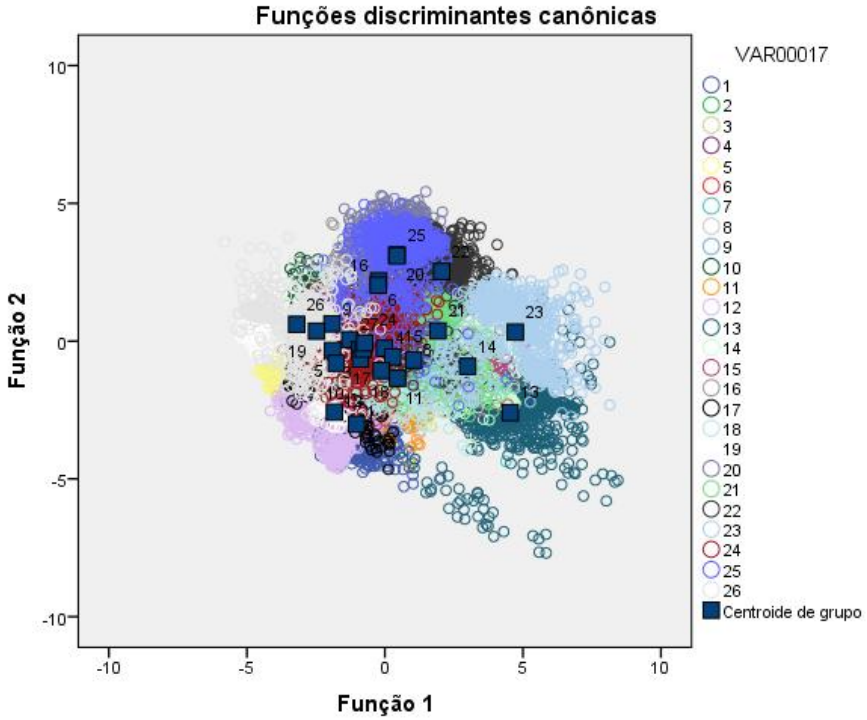


FIGURA 40 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO LETTER
FONTE: O autor (2015)

O método do *k*-segmentos apresentou desempenho melhor que o dos centroides, com 83,2% de eficiência (apêndices 4, 5 e 6), e individualmente nenhuma classe teve resultado inferior a 62%. As classes 1, 4, 8, 10, 12, 13, 21, 22 e 23 obtiveram desempenho superior a 89%. Na figura 41, observa-se que as CPs formam um ajuste que não é o ideal para o ótimo desempenho desse método, devido à proximidade e o cruzamento das curvas, porém o método *k*-segmentos obteve um bom desempenho na classificação.

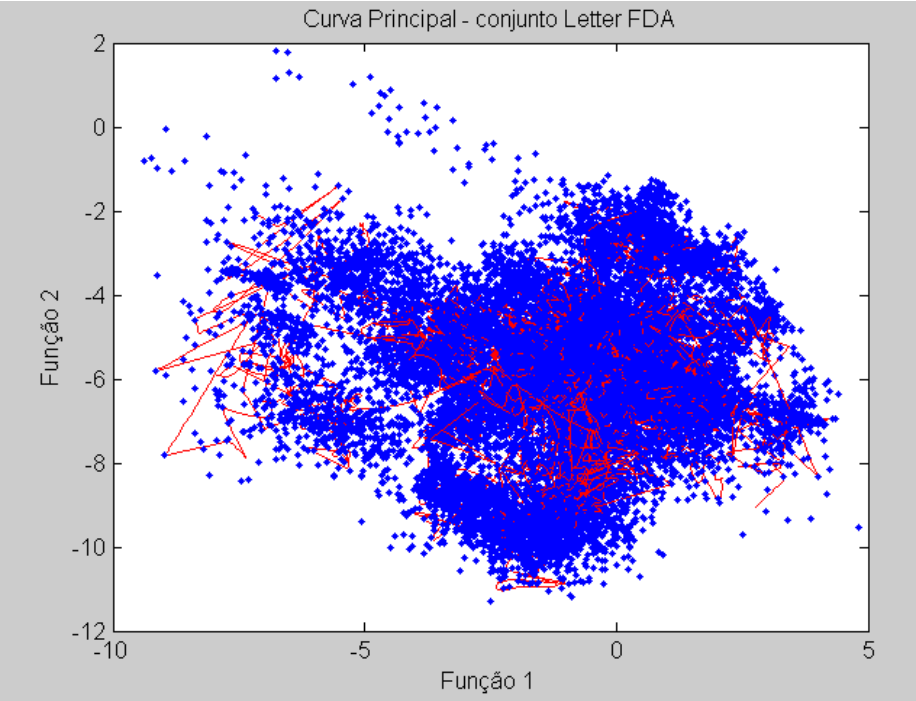


FIGURA 41 - CURVA PRINCIPAL PARA O CONJUNTO LETTER
FONTE: O autor (2015)

4.2.13 Balance

O conjunto *balance-scale* trata-se de um banco de dados para um modelo psicológico (psicologia cognitiva). Foi desenvolvido por Siegler (1976) e possui 625 observações, 4 variáveis e 3 classes. Para o método dos centroides, a matriz de confusão (TABELA 16) mostra que houve erro somente com a classificação de vetores das classes 2 e 3 na classe 1. A explicação para este resultado é que a classe 1 encontra-se no 'meio' das duas classes, com o seu centroide próximo aos outros dois (FIGURA 42). O método dos *k*-segmentos foi muito superior na classificação, com apenas 56 erros contra 192 erros do método dos centroides. A percentagem de acertos mostra a superioridade do método *k*-segmentos com 91,04% contra 69,3%.

TABELA 15 - MATRIZ DE CONFUSÃO PARA O CONJUNTO *BALANCE* PARA FDA CENTROIDE X FDA *K*-SEGMENTOS

FDA centroides					**FDA *k*-segmentos - *k* =5**		
Classe	**Classe predita**				**Classe predita**		
	Tamanho da classe	**1**	**2**	**3**	**1**	**2**	**3**
1	49	**49**	0	0	**45**	2	2
		100%	0,00%	0,00%	**91,84%**	4,08%	4,08%
2	288	96	**192**	0	22	**262**	4
		33,3%	**66,67%**	0,00%	7,64%	**90,97%**	1,39%
3	288	96	0	**192**	22	4	**262**
		3,33%	0,00%	**66,67%**	7,64%	1,39%	**90,97%**
% de acertos:		**69,3%**			**% de acertos:**		**91.04%**

FONTE: O autor (2015)

O gráfico dos centroides e das curvas principais (FIGURAS 42 e 43) apresentou *design* diferente dos gráficos anteriores e, mesmo entre os dois, o desenho é diferente. O método *k*-segmentos foi mais eficiente ao evitar a interseção dos escores.

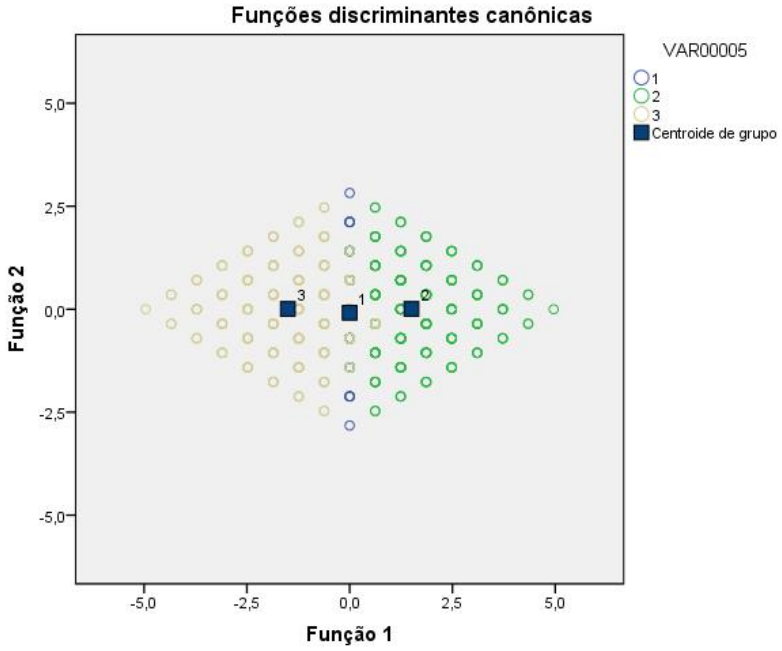


FIGURA 42 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO BALANCE
FONTE: O autor (2015)

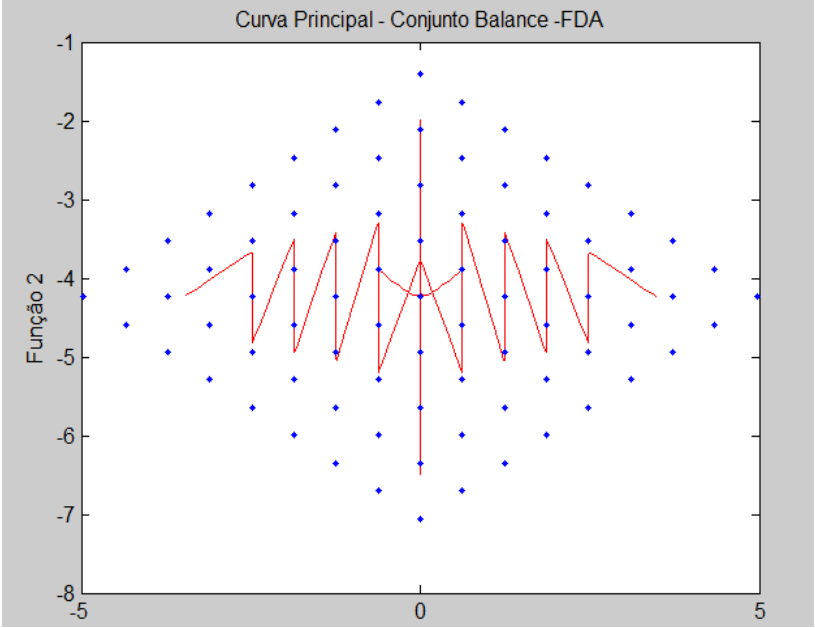


FIGURA 43 - CURVA PRINCIPAL PARA O CONJUNTO BALANCE
FONTE: O autor (2015)

4.2.14 Veículo

Conjunto com 846 observações, 18 variáveis e 4 classes, construído por JP Siebert e tem por objetivo a classificação de 4 tipos de veículos, através das características externas (variáveis). Na análise do gráfico dos centroides (FIGURA 44), pode-se observar que as classes 1 e 2 foram muito bem discriminadas, com os centroides no centro dos respectivos conjuntos e com pouca interseção de seus escores com as demais classes, gerando expectativa de um excelente resultado para estas duas classes. Esta esperança se pode confirmar ao analisar a matriz de confusão (TABELA 17), com 97,7% e 94,5% de acerto para as classes 1 e 2. Já as classes 3 e 4 apresentaram forte interseção dos escores e com centroides muito próximos, fazendo com que fossem esperados resultados ruins para estas classes, que foram confirmados na matriz de confusão, a qual apresentou apenas 57,1% e 60,4% de acerto para as classes 3 e 4 (resultado regular).

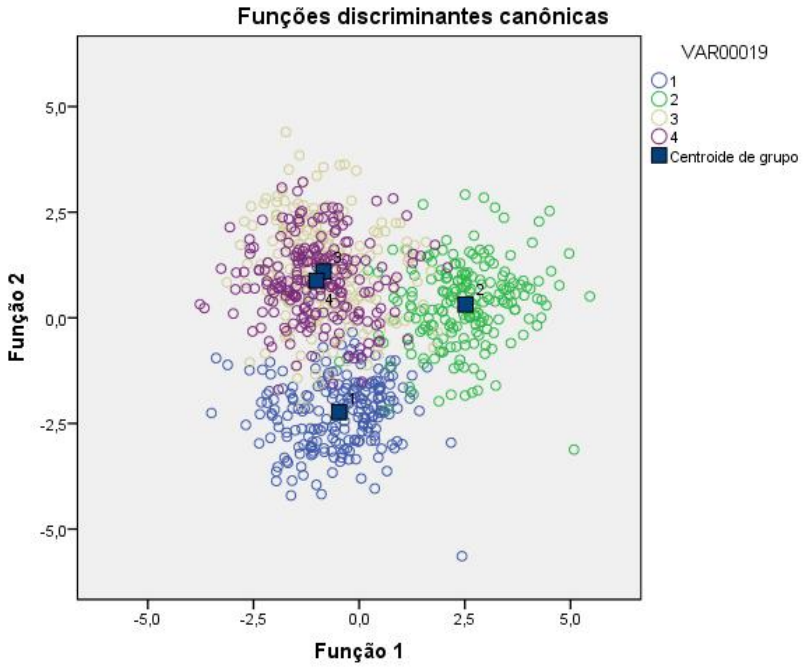


FIGURA 44 - GRÁFICO DAS FUNÇÕES DISCRIMINANTES PARA O CONJUNTO VEÍCULO
FONTE: O autor (2015)

TABELA 16 - MATRIZ DE CONFUSÃO PARA O CONJUNTO VEÍCULO PARA FDA CENTROIDE X FDA *K*-SEGMENTOS

FDA centroides						**FDA *k*-segmentos - *k* =4**			
Classe predita						**Classe predita**			
Classe	**Tamanho da classe**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1	218	**213**	1	0	4	**203**	3	10	2
		97,7%	0,5%	0,0%	1,8%	**93,1%**	1,4%	4,6%	0,9%
2	199	6	**188**	4	1	1	**190**	3	5
		3,0%	**94,5%**	2,0%	0,5%	0,5%	**95,4%**	1,5%	2,5%
3	217	14	15	**124**	64	11	8	**131**	67
		6,5%	6,9%	**57,1%**	29,5%	5,1%	3,7%	**60,3%**	30,8%
4	212	11	8	65	**128**	18	6	58	**130**
		5,20%	3,8%	30,7%	**60,4%**	8,5%	2,8%	27,3%	**61,3%**
% de acertos:	**77,2%**					**% de acertos:**		**77,3%**	

FONTE: O autor (2015)

O gráfico das curvas principais apresentam a separação em duas nuvens de escores, das classes 1 e 2, e 3 e 4. Devido a interseção da curva e de uma ajuste não linear, o desempenho do método *k*-segmentos foi prejudicado (FIGURA 45).

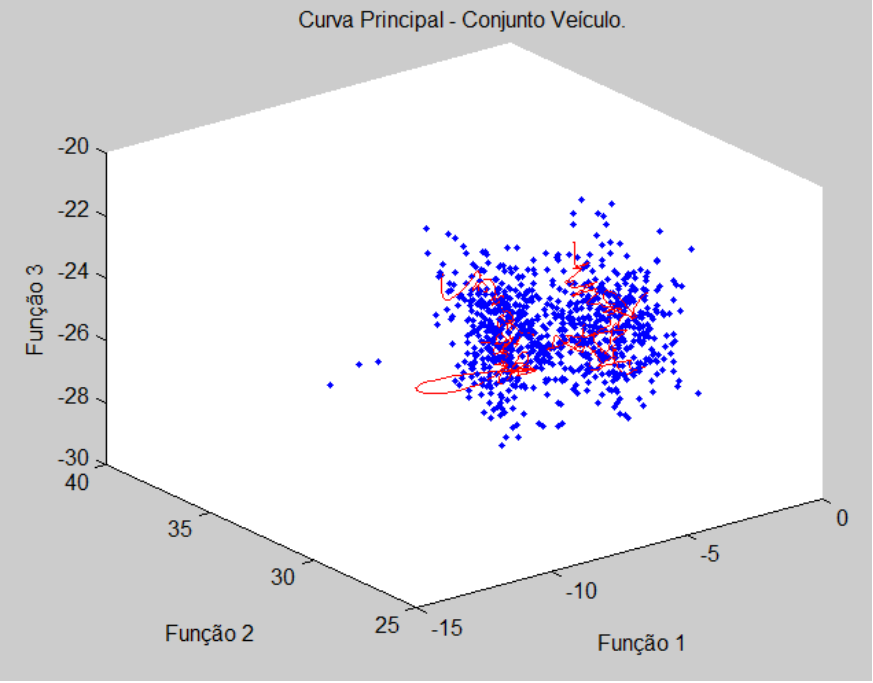


FIGURA 45 - CURVA PRINCIPAL PARA O CONJUNTO VEÍCULO
FONTE: O autor (2015)

4.3 LFDA-CENTROIDES *VERSUS* LFDA K-SEGMENTOS

Nesta seção é comparada a eficiência do método dos centroides com o método *k*-segmentos utilizando a técnica LFDA. A comparação é realizada para todos os conjuntos trabalhados na FDA.

4.3.1 Iris

O resultado apresentado pela matriz de confusão (TABELA 18) mostra o bom desempenho das duas técnicas, mas a técnica LFDA com centroides obteve o pior resultado, com 92,66%, Já o método *k*-segmentos apresentou resultado igual ao obtido na FDA (98,66%). Convém destacar que quando são utilizadas as CPs tem-se melhor eficiência que a dos centroides. Com apenas dois erros a menos (98 erros de classificação contra 100) o método que utiliza as CPs foi ligeiramente melhor (tanto para FDA como para LFDA).

TABELA 17 - MATRIZ DE CONFUSÃO PARA O CONJUNTO IRIS PARA LFDA CENTROIDES X LFDA *K*-SEGMENTOS

LFDA centroide					**LFDA *k*-segmentos *k=3***		
Classe	**Tamanho da Classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	50	**50**	0	0	**50**	0	0
		100,00%	0,00%	0,00%	**100,00%**	0,00%	0,00%
2	50	0	**46**	4	0	**48**	2
		0,00%	**92,00%**	8,00%	0,00%	**96,00%**	4,00%
3	50	0	7	**43**	0	0	**50**
		0,00%	14,00%	**86,00%**	0,00%	0,00%	**100,00%**
Casos classificados corretamente: 92,66%							**98,66%**

FONTE: O autor (2015)

Na análise da figura 46 é fácil ver as curvas principais bem definidas, ‘passando’ pelo centro das classes, o que favorece o desempenho dessa técnica, confirmando o ótimo desempenho na classificação.

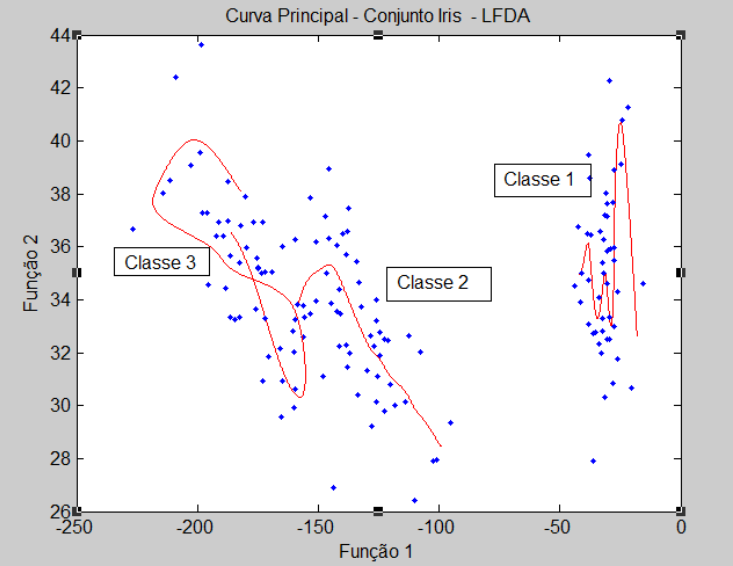


FIGURA 46 - CURVA PRINCIPAL PARA O CONJUNTO IRIS - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.2 Medical

As duas técnicas FDA e LFDA por centroides apresentaram o mesmo resultado geral na classificação, porém na comparação com a técnica que utiliza as CPs o desempenho foi inferior. Tanto para FDA como para LFDA, o resultado apresentado, quando utilizado o algoritmo *k*-segmentos é muito superior, com 92,5% (TABELA 5) e 92,40% (TABELA 19), respectivamente. A figura 47 mostra as curvas principais para cada classe, bem discriminadas (com pouca interseção), passando pelo meio de suas respectivas classes, o que favorece o bom desempenho do algoritmo *k*-segmentos. A classe melhor discriminada (classe 3) obteve 100% de eficiência utilizando o algoritmo *k*-segmentos nas duas técnicas, FDA e LFDA.

TABELA 18 - MATRIZ DE CONFUSÃO PARA O CONJUNTO *MEDICAL* PARA LFDA CENTROIDES X LFDA *K*-SEGMENTOS

		LFDA- centroide			**LFDA *k*-segmentos *k*=2**		
		Classe predita			**Classe predita**		
Classe	**Tamanho da Classe**	**1**	**2**	**3**	**1**	**2**	**3**
1	26		3	1	**24**	2	0
		84,62%	11,54%	3,85%	**92,31%**	7,69%	0,00%
2	18	0	**17**	1	1	**16**	1
		0,00%	**94,44%**	5,56%	5,56%	**88,89%**	5,56%
3	10	0	1	**9**	0	1	**9**
		0,00%	10,00%	**90,00%**	0,00%	11,11%	**100,00%**
Classificação corretas			**90,55%**				**92,40%**

FONTE: O autor (2015)

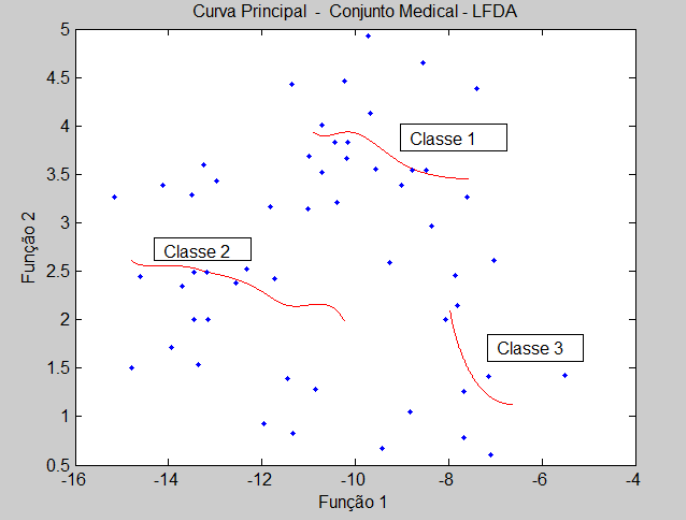


FIGURA 47 - CURVA PRINCIPAL PARA O CONJUNTO MEDICAL -TÉCNICA LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.3 Wave

A matriz de confusão (TABELA 20) mostra o melhor desempenho do algoritmo *k*-segmentos sobre LFDA com centroides, que não é superior ao obtido pela FDA com CP. Pode-se observar na comparação das tabelas 6 e 19, que a LFDA, cuja proposta é de ser mais eficiente que a FDA, não foi superior a mesma.

Na classificação individual, a primeira classe apresentou resultados bem diferentes na utilização de centroide e CP, com resultado muito ruim para com centroides (49,25% de acertos). Nas demais classes foi superior aos demais métodos.

TABELA 19 - MATRIZ DE CONFUSÃO PARA O CONJUNTO WAVE PARA LFDA CENTROIDES X K-SEGMENTOS

		LFDA centroide			**LFDA *k*-segmentos k=5**		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	1657	**816**	427	414	**1355**	181	121
		49,25%	25,77%	24,98%	**81,77%**	10,92%	7,30%
2	1647	3	**1566**	78	98	**1435**	114
		0,18%	**95,08%**	4,74%	5,95%	**87,13%**	6,92%
3	1696	2	74	**1620**	141	56	**1499**
		0,12%	4,36%	**95,52%**	8,31%	3,30%	**88,38%**
Casos classificados corretamente: 80,04%					**Casos classificados corretamente: 85,78%**		

FONTE: O autor (2015)

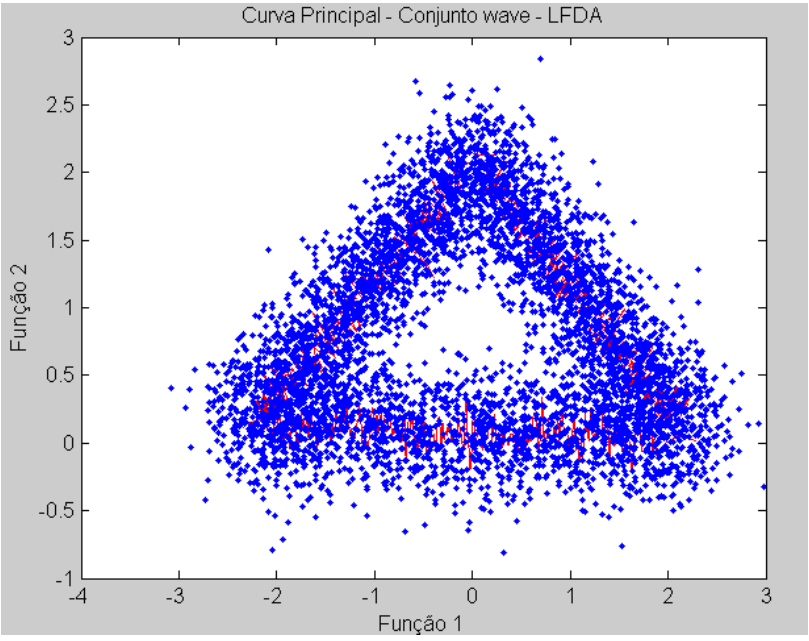


FIGURA 48 - CURVA PRINCIPAL PARA O CONJUNTO WAVE - LFDA K-SEGMENTOS
FONTE: O autor (2015)

Na figura 48, o gráfico das curvas sobre as classes é semelhante ao obtido nos dois métodos que utilizam a CP. De modo geral, o algoritmo *k*-segmentos apresentou bom resultado, tanto para LFDA, como para FDA.

4.3.4 Álcool

Neste conjunto a LFDA também apresentou resultado muito inferior às demais técnicas, com resultado apenas regular (68,83% de acertos), resultado este apresentado na tabela 21. As técnicas FDA e LFDA quando utilizam o algoritmo *k*-segmentos apresentam resultado superior ao dos centroides, com 89,61% e 84,41%, respectivamente, contra 80,52% e 68,83% de acertos por centroides. Devido à interseção dos escores das classes 2 e 3 (FIGURA 49) que afeta a eficiência na classificação, o resultado individual para todos os métodos aplicados foi razoável.

TABELA 20 - MATRIZ DE CONFUSÃO PARA O CONJUNTO ÁLCOOL PARA LFDA X *K*-SEGMENTOS

FDA centroide					**LFDA *k*-segmentos k=4**		
Classe	**Tamanho da Classe**	**Classe predita**			**Classe predita**		
		Kirsch	***Mirab***	***Poire***	***Kirsch***	***Mirab***	***Poire***
Kirsch	18	**18**	0	0	**18**	0	0
		100,00%	0,00%	0,00%	**100,00%**	0,00%	0,00%
Mirab	29	1	**17**	11	0	**23**	6
		3,45%	**58,62%**	37,93%	0,00%	**79,31%**	17,24%
Poire	30	1	11	**18**	0	6	**24**
		3,33%	36,67%	**60,00%**	0,00%	20,00%	**80,00%**
Casos classificados corretamente: 68,83%					**84,41%**		

FONTE: O autor (2015)

O gráfico das curvas principais (FIGURA 49) para LFDA apresenta a interseção das curvas nas classes 2 e 3, fato que afeta o desempenho do algoritmo *k*-segmentos, diferentemente da classe que tem os escores discriminantes separados das demais classes com a CP sem interseção, o que é bom para o método *k*-segmentos.

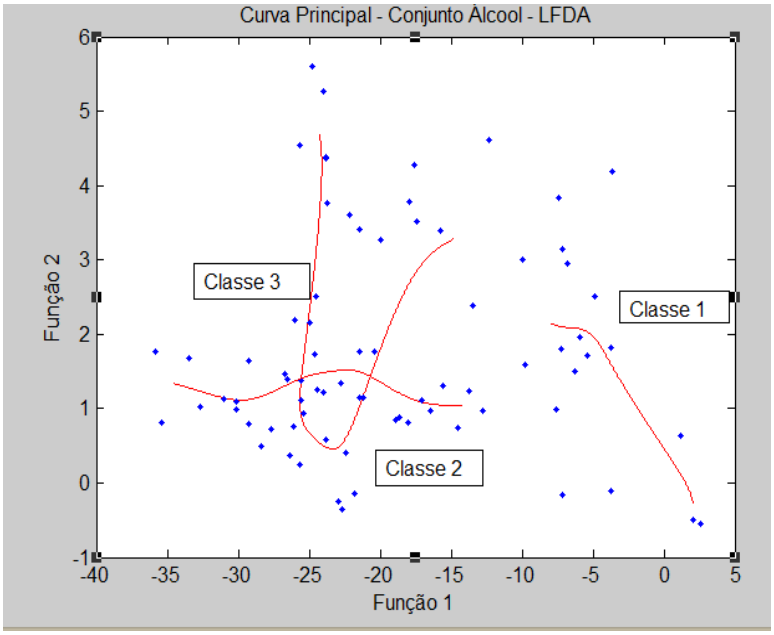


FIGURA 49 - CURVA PRINCIPAL PARA O CONJUNTO ÁLCOOL - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.5 Tiroide

A técnica *k*-segmentos para FDA aplicada a este conjunto obteve melhor resultado que as demais e com pequena vantagem em relação à obtida pelo *k*-segmentos para LFDA (97,21% contra 96,74%), conforme é apresentado nas tabelas 9 e 22. A LFDA com centroides é a técnica com pior desempenho, mas ficou próximo ao resultado da FDA também por centroides (93,49% contra 94,42%). Do mesmo modo que ocorre na FDA, a interseção dos escores interferiu no desempenho das duas técnicas: Na LFDA com centroides há interseção das classes 2 e 3 com a classe 1; para o algoritmo *k*-segmentos há interseção da classe 1 com a 2 e 3. As duas técnicas que utilizam as CPs foram eficientes.

TABELA 21 - MATRIZ DE CONFUSÃO PARA O CONJUNTO TIROIDE PARA LFDA X *K*-SEGMENTOS

		LFDA centroide			**LFDA k-segmentos *k*=4**		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	150	**150**	0	0	**147**	2	1
		100,00%	0,00%	0,00%	**98,00%**	1,33%	0,67%
2	35	8	**27**	0	1	**34**	0
		22,86%	**77,14%**	0,00%	2,86%	**97,14%**	0,00%
3	30	6	0	**24**	3	0	**27**
		20,00%	0,00%	**80,00%**	10,00%	0,00%	**90,00%**
Casos classificados corretamente: 93,49%						**96,74%**	

FONTE: O autor (2015)

Na análise do gráfico das curvas principais (FIGURA 50), o bom desempenho da técnica *k*-segmentos deve-se ao fato das curvas 'passarem' no meio das classes e de não haver interseção entre as CPs. Pode-se destacar que a discriminação das classes tanto pela FDA como pela LFDA foi eficiente, embora haja uma pequena interseção entre os escores das classes.

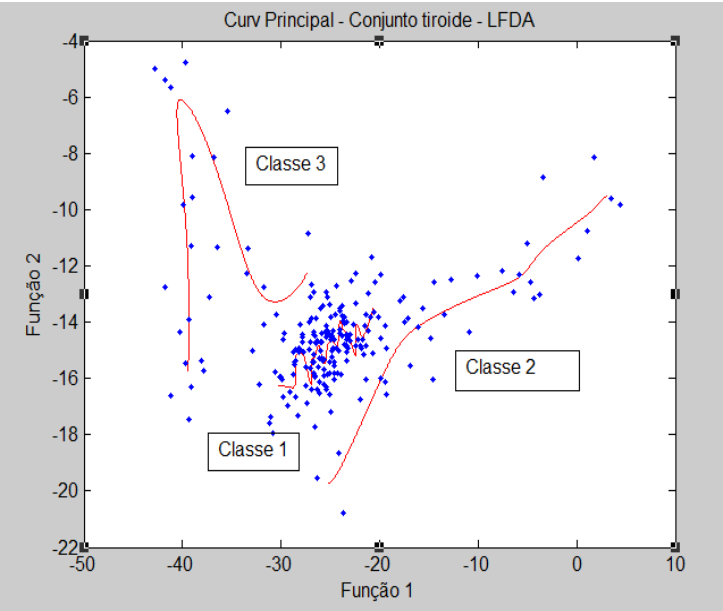


FIGURA 50 - CURVA PRINCIPAL PARA O CONJUNTO TIROIDE - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.6 Glass

A matriz de confusão do conjunto Glass (TABELA 23), mostra os resultados obtidos pela LFDA. O algoritmo *k*-segmentos apresentou melhor resultado que todas as demais técnicas. Já a LFDA por centroides novamente apresentou o pior resultado, com 48,59% de acertos. A classe 2 apresentou o pior resultado individual (quando utilizados os centroides), que comprometeu bastante o resultado geral da LFDA por centroides.

TABELA 22 - MATRIZ DE CONFUSÃO PARA O CONJUNTO *GLASS* PARA LFDA X *K*-SEGMENTOS

LFDA centroide						**LFDA k-segmentos k=5**					
Classe predita						**Classe predita**					
1	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
39	8	23	0	0	0	**53**	9	8	0	0	0
55,71 %	11,43 %	32,86 %	0,0%	0,0%	0,00 %	**75,71 %**	12,86 %	11,43 %	0,0%	0,00 %	0,00 %
36	**16**	10	3	10	1	16	**51**	6	0	2	1
47,37 %	**21,05 %**	13,16 %	3,95 %	13,16 %	1,32 %	21,05 %	**67,11 %**	7,89 %	0,00%	2,63 %	1,32 %
5	4	**8**	0	0	0	5	3	**9**	0	0	0
29,41 %	23,53 %	**47,06 %**	0,00 %	0,00%	0,00 %	29,41 %	17,65 %	**52,94 %**	0,00%	0,00 %	0,00 %
0	0	0	**9**	3	1	0	0	0	**13**	0	0
0,00 %	0,00 %	0,00 %	**69,23 %**	23,08 %	7,69 %	0,00 %	0,00 %	0,00 %	**100,00 %**	0,00 %	0,00 %
0	0	0	0	**9**	0	0	1	0	0	**8**	0
0,00 %	0,00 %	0,00 %	0,00 %	**100,00 %**	0,00 %	0,00 %	11,11 %	0,00 %	0,00%	**88,89 %**	0,00 %
1	1	1	1	2	**23**	1	0	0	0	0	**28**
3,45 %	3,45 %	3,45 %	3,45 %	6,90%	**79,31 %**	3,45 %	0,00 %	0,00 %	0,00%	0,00 %	**96,55 %**
Casos classificados corretamente: 48,59%						**Casos classificados corretamente: 75,7%**					

FONTE: O autor (2015)

O gráfico das CPs (FIGURA 51) apresenta as curvas principais para as classes com interseção, curvas próximas uma das outras e também a interseção dos escores entre as classes. Este fato corrobora com o resultado apresentado na matriz

de confusão, com resultado apenas razoável em termos de eficiência de classificação.

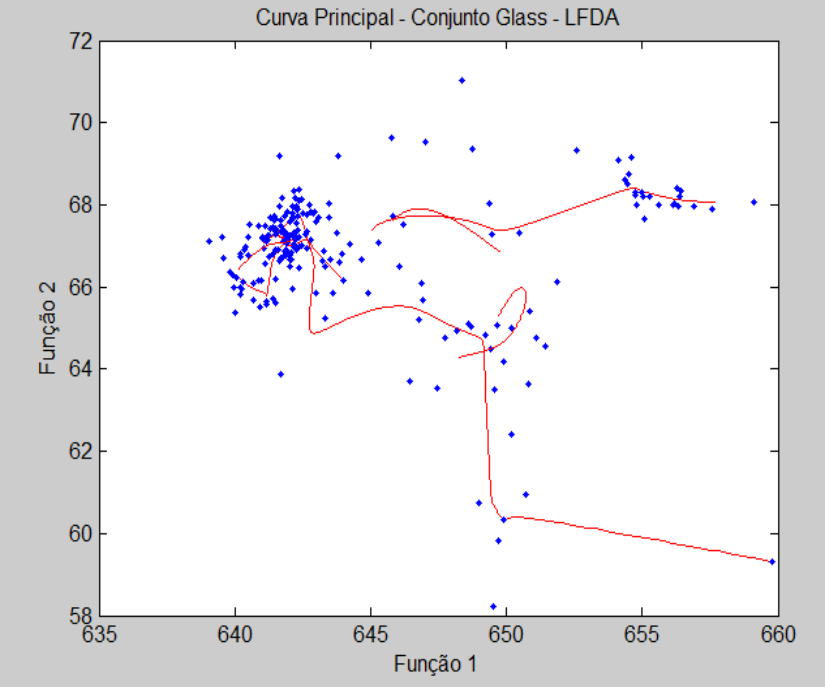


FIGURA 51 - CURVA PRINCIPAL PARA O CONJUNTO GLASS - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.7 Futebol

A matriz de confusão (TABELA 24) apresenta o resultado regular das duas técnicas comparadas nesta seção. Resultado bastante afetado pela interseção dos escorres das classes 2 e 3. Individualmente, a classe 3 apresentou o pior resultado para LFDA, com apenas 46,66% de acerto. Comparando o resultado da LFDA com a FDA, a primeira obteve resultado inferior à segunda. Também a LFDA, quando utilizou a linha poligonal, apresentou pior resultado se comparado com a FDA por CP. A técnica por *k*-segmentos apresentou resultado ruim na 2ª classe, com apenas 50% de classificação correta.

TABELA 23 - MATRIZ DE CONFUSÃO PARA O CONJUNTO FUTEBOL PARA LFDA X *K*-SEGMENTOS

		LFDA centroide			**LFDA *k*-segmentos *k=3***		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	150	**27**	1	2	**27**	0	3
		90,00%	3,33%	6,66%	**90,00%**	0,00%	10,00%
2	35	2	**21**	7	1	**15**	14
		6,66%	**70,00%**	23,33%	3,33%	**50,00%**	46,67%
3	30	4	12	**14**	0	10	**20**
		13,33%	40,00%	**46,66%**	0,00%	33,33%	**66,67%**
Casos classificados corretamente: 68,88%					**68,88%**		

FONTE: O autor (2015)

O gráfico das curvas principais (FIGURA 52), mostra grande interseção das classes 2 e 3, com as curvas principais muito próximas, o que prejudica a eficiência do método *k*-segmentos, pois devido à variabilidade própria de cada classe, um vetor pertencente a classe 2 pode estar mais próximo da linha poligonal da classe 3 e deste modo interferir no cálculo das distâncias do vetor a ser classificado, categorizando-o como da classe 3, sendo da classe 2.

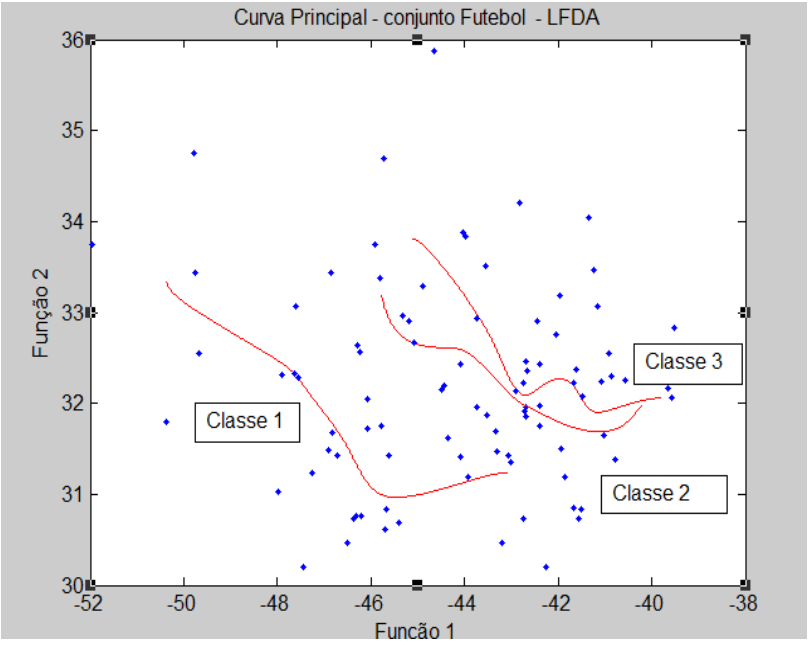


FIGURA 52 - CURVA PRINCIPAL PARA O CONJUNTO FUTEBOL - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.8 Besouro

Da mesma maneira que ocorre na FDA, repete-se para LFDA. Apresenta-se excelente resultado, tanto para LFDA com centroides, como para com CP. Na Tabela 25, é fácil ver que em todas as classes e no resultado geral apresentam 100% de eficiência.

TABELA 24 - MATRIZ DE CONFUSÃO PARA O CONJUNTO BESOURO PARA LFDA CENTROIDES X *K*-SEGMENTOS

		LFDA centroide			**LFDA *k*-segmento *k=2***		
		Classe predita			**Classe predita**		
Classe	**Tamanho da classe**	**1**	**2**	**3**	**1**	**2**	**3**
1	21	**26**	0	0	**26**	0	0
		100,00%	0,00%	0,00%	**100,00%**	0,00%	0,00%
2	22	0	**18**	0	0	**18**	0
		0,00%	**100,00%**	0,00%	0,00%	**100,00%**	0,00%
3	31	0	0	**10**	0	0	**10**
		0,00%	0,00%	**100,00%**	0,00%	0,00%	**100,00%**
				% de acerto: 100%			**% de acerto: 100%**

FONTE: O autor (2015)

Na análise das curvas principais, na figura 53, é fácil ver que as classes estão bem discriminadas (sem interseção entre classes), com as curvas no meio de suas respectivas classes.

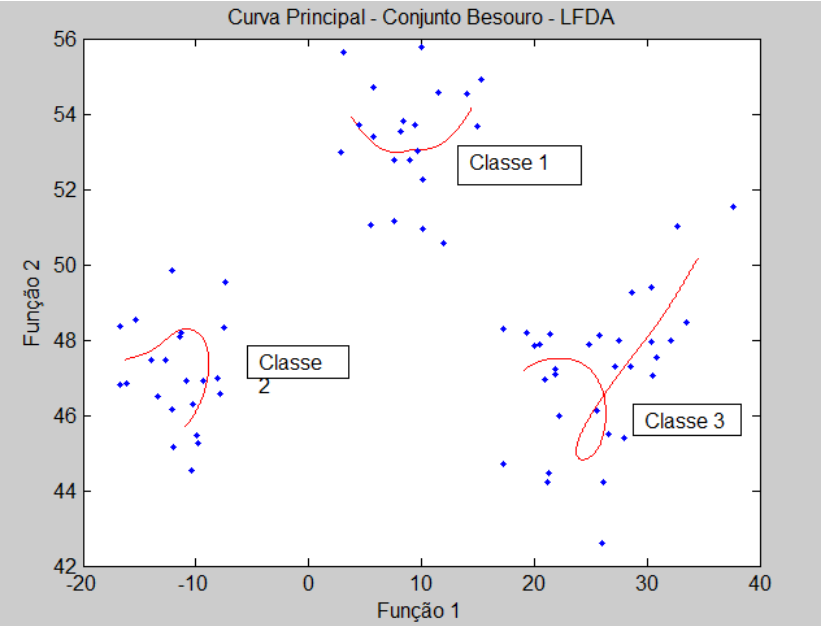


FIGURA 53 - CURVAS PRINCIPAIS PARA O CONJUNTO BESOURO - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.9 Letter

A LFDA com o algoritmo *k*-segmentos apresentou desempenho muito ruim, com apenas 24,40% de classificações corretas. Este resultado é apresentado nos apêndices 7, 8 e 9. Apenas as variáveis 23, 24 e 26 obtiveram desempenho superior a 50%. Este é o pior desempenho de qualquer uma das técnicas utilizadas neste trabalho. Os apêndices 10, 11 e 12 apresentam o resultado obtido pelo uso da LFDA com o algoritmo *k*-segmentos.

O gráfico das curvas principais (FIGURA 54) apresentou o conjunto de escorres mais concentrados para a FDA e consequentemente afetou bastante a eficiência da técnica LFDA.

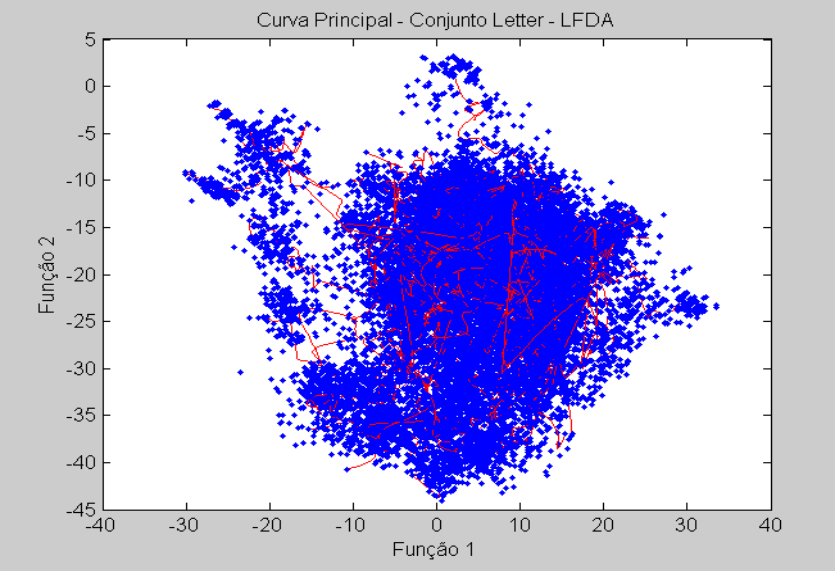


FIGURA 54 - CURVA PRINCIPAL PARA O CONJUNTO LETTER - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.10 Balance

A matriz de confusão apresenta o bom desempenho da técnica LFDA por *k*-segmentos, com 92,48%, que é o mesmo resultado obtido na FDA. Este resultado também é semelhante ao obtido individualmente por classe (TABELA 26). Na LFDA por centroides o desempenho individual é semelhante ao dado pela FDA pela mesma técnica, mas muito inferior ao obtido pelo algoritmo *k*-segmentos. Apenas a classe 1 apresentou desempenho excelente, com 100% de acertos. O destaque é que o algoritmo *k*-segmentos é muito superior ao dos centroides, 68,64% contra 92,48%, semelhante ao ocorrido com a FDA.

TABELA 25 - MATRIZ DE CONFUSÃO PARA O CONJUNTO *BALANCE* PARA LFDA X *K*-SEGMENTOS

		LFDA centroide			**LFDA *k*-segmentos *k=2***		
		Classe predita			**Classe predita**		
Classe	**Tamanho da classe**	**1**	**2**	**3**	**1**	**2**	**3**
1	49	**49**	0	0	**49**	0	0
		100,00%	0,00%	0,00%	**100,00%**	0,00%	0,00%
2	288	98	**190**	0	23	**241**	24
		34,03%	**65,97%**	0,00%	7,99%	**83,68%**	8,33%
3	288	98	0	**190**	0	0	**288**
		34,03%	0,00%	**65,97%**	0,00%	0,00%	**100,00%**
			% de acerto: 68,64%			**% de acerto: 92,48%**	

FONTE: O autor (2015)

As curvas principais de cada classe sobre os escores discriminantes são dadas na figura 55. As curvas principais 'passam' pelo meio de cada classe, o que favorece o bom desempenho do algoritmo *k*-segmentos.

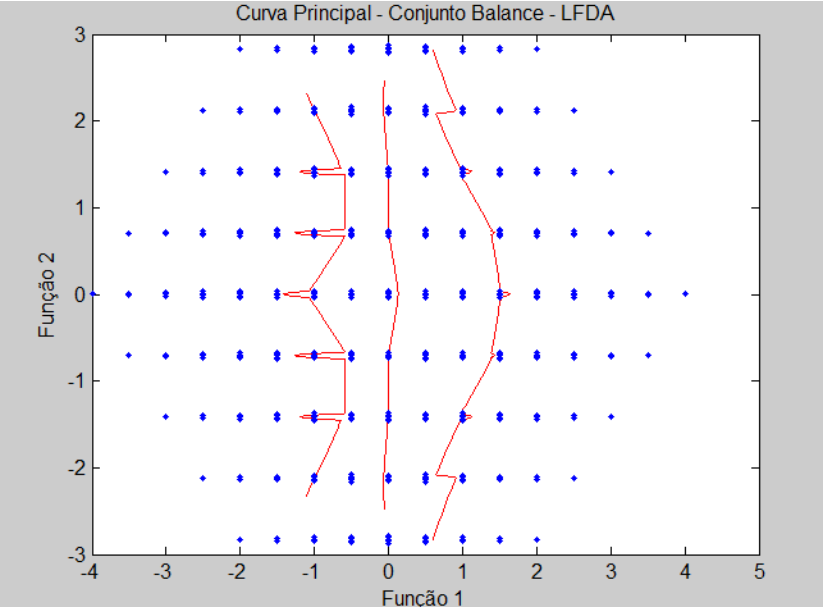


FIGURA 55 - CURVA PRINCIPAL PARA O CONJUNTO BALANCE - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.11 Abalone

Com resultados próximos, a LFDA por centroides e *k*-segmentos (TABELA 27) não se distanciou neste conjunto das técnicas FDA por centroides ou por k-segmentos (TABELA 15). A percentagem de acertos é de 48,47% contra 54,32% da FDA (ambas com centroides), como apresentado na página 91. O melhor desempenho é da FDA por centroides, pois utilizando *k*-segmentos, apenas a LFDA apresentou resultado próximo (50,05%). Na análise individual, a classe 1 apresentou desempenho muito ruim para LFDA com centroides e com *k*-segmentos (19,31% e 24,28%, respectivamente) e em segundo lugar a 2ª classe com menos de 50% de acertos para as duas técnicas.

TABELA 26 - MATRIZ DE CONFUSÃO PARA O CONJUNTO ABALONE - LFDA *K*-SEGMENTOS

		LFDA centroide			**LFDA *k*-segmentos *k*=2**		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	1328	**295**	656	577	**371**	645	512
		19,31%	42,93%	37,76%	**24,28%**	42,21%	33,51%
2	1307	248	**635**	424	326	**603**	378
		18,97%	**48,58%**	32,44%	24,94%	**46,14%**	28,92%
3	1342	92	155	**1095**	135	90	**1117**
		6,86%	11,55%	**81,59%**	10,06%	6,71%	**83,23%**
Casos classificados corretamente: 48,47%					**50,05%**		

FONTE: O autor (2015)

A figura 56 mostra o gráfico das curvas principais para o conjunto abalone com forma semelhante ao obtido na FDA, porém com pontos discriminantes, formando linhas paralelas que separam os dados, pois a LFDA não efetuou de forma satisfatória a discriminação das classes. O cruzamento e sobreposição de curvas principais e o fato de que a transformação dos dados pela LFDA não separou as classes explicam o desempenho ruim das técnicas utilizadas para este conjunto.

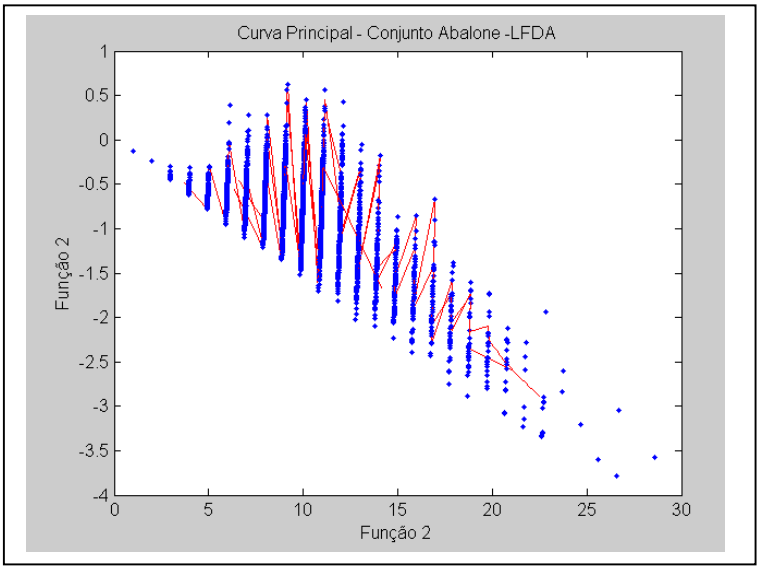


FIGURA 56 - CURVA PRINCIPAL PARA O CONJUNTO ABALONE - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.12 Segment

As duas técnicas apresentaram resultado pouco diferente, com melhor resultado para a LFDA com *k*-segmentos, com 78,78% e 82,68%, respectivamente (TABELAS 28 e 29). Porém, esse resultado é muito inferior ao resultado obtido pela FDA com centroides e com *k*-segmentos, com resultados de 91,77% e 92,55%. A LFDA por centroides apresentou resultado inferior das demais, com 78,78% de acertos. Individualmente as classes 2, 5 e 7 apresentaram resultado inferior a 66% e a melhor eficiência foi para a classe 3, com 100%, o que é bom para o modelo utilizado.

TABELA 27 - MATRIZ DE CONFUSÃO PARA O CONJUNTO *SEGMENT* PARA LFDA CENTROIDES

LFDA centroides								
Classe	**Tamanho**	**Classe predita**						
	da classe	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	330	**298**	1	0	0	29	0	2
		90,30%	0,30%	0,00%	0,00%	8,79%	0,00%	0,61%
2	330	0	**216**	0	0	42	14	58
		0,00%	**65,45%**	0,00%	0,00%	12,73%	4,24%	17,58%
3	330	0	0	**330**	0	0	0	0
		0,00%	0,00%	**100,00%**	0,00%	0,00%	0,00%	0,00%
4	330	0	2	0	**327**	0	0	1
		0,00%	0,61%	0,00%	**99,09%**	0,00%	0,00%	0,30%
5	330	0	79	0	0	**168**	15	68
		0,00%	23,94%	0,00%	0,00%	**50,91%**	4,55%	20,61%
6	330	15	6	0	0	29	**264**	16
		4,55%	1,82%	0,00%	0,00%	8,79%	**80,00%**	4,85%
7	330	0	56	0	0	37	20	**217**
		0,00%	16,97%	0,00%	0,00%	11,21%	6,06%	**65,76%**
		% de acertos:		**78,78%**				

FONTE: O autor (2015)

A matriz de confusão da LFDA com *k*-segmentos (TABELA 29) apresentou resultado mais próximo dos resultados obtidos pela FDA com centroides e com *k*-segmentos. Apenas a classe 7 obteve eficiência inferior a 70% de acertos. Embora a percentagem de acertos seja razoável para este conjunto, ainda é inferior ao obtido pela FDA por centroides e por *k*-segmentos.

TABELA 28 - MATRIZ DE CONFUSÃO PARA O CONJUNTO *SEGMENT* PARA *K*-SEGMENTOS

		LFDA *k*-segmentos		***k=5***				
Classe	**Tamanho da classe**	**Classe predita**						
		1	**2**	**3**	**4**	**5**	**6**	**7**
1	330	**286**	2	0	0	0	36	6
		86,67%	0,61%	0,00%	0,00%	0,00%	10,91%	1,82%
2	330	1	**237**	0	0	30	18	44
		0,30%	**71,82%**	0,00%	0,00%	9,09%	5,45%	13,33%
3	330	0	0	**330**	0	0	0	0
		0,00%	0,00%	**100,0%**	0,00%	0,00%	0,00%	0,00%

continua

Classe	Tamanho da classe	LFDA *k*-segmentos *k=5*						
		Classe predita						
		1	2	3	4	5	6	7
								conclusão
4	330	0	0	0	**327**	1	0	2
		0,00%	0,00%	0,00%	**99,09%**	0,30%	0,00%	0,61%
5	330	0	9	0	0	**296**	6	19
		0,00%	2,73%	0,00%	0,00%	**89,70%**	1,82%	5,76%
6	330	29	3	0	1	5	**253**	39
		8,79%	0,91%	0,00%	0,30%	1,52%	**76,67%**	11,82%
7	330	0	72	0	1	60	16	**181**
		0,00%	21,82%	0,00%	0,30%	18,18%	4,85%	**54,85%**
% de acertos								**82,68%**

FONTE: O autor (2015)

Na análise do gráfico das linhas poligonais sobre os escores discriminantes (FIGURA 57), é fácil ver que o mesmo é semelhante ao obtido na FDA *k*-segmentos, mas as linhas neste gráfico estão mais concentradas e com bastante interseção entre as mesmas, o que prejudicou o desempenho do modelo.

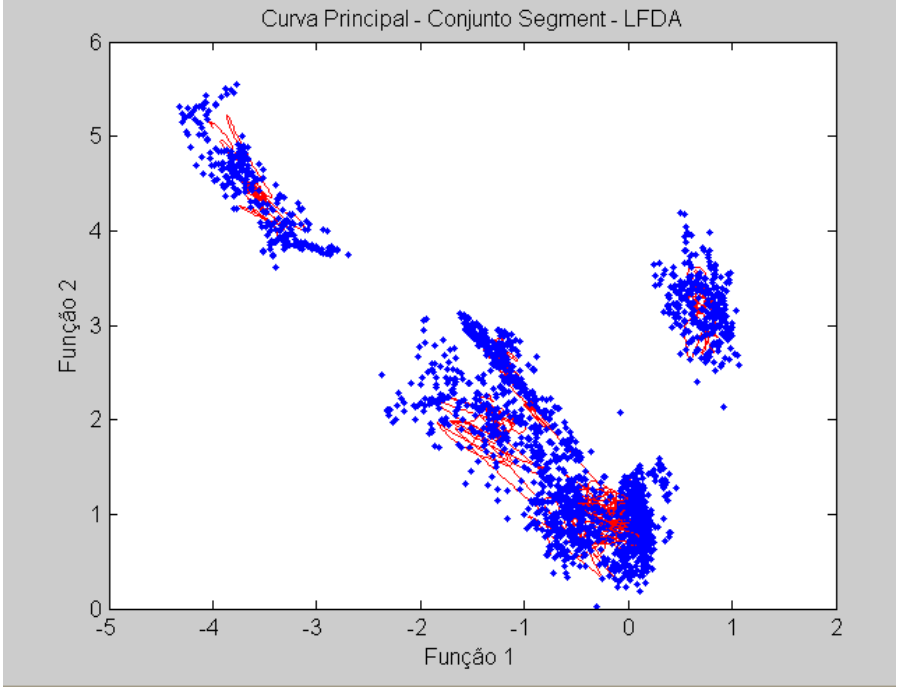


FIGURA 57 - CURVA PRINCIPAL PARA O CONJUNTO SEGMENT - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.13 Wine

A LFDA por centroides e pelo algoritmo *k*-segmentos apresentam 73,03% e 97,75% de eficiência, respectivamente, conforme é apresentado na tabela 30. As duas técnicas apresentaram resultado inferior ao obtido pela FDA, com destaque para a LFDA por centroides, que obteve resultado muito distante dos 100% de classificação correta.

TABELA 29 - MATRIZ DE CONFUSÃO PARA O CONJUNTO WINE PARA FDA CENTROIDES X FDA K-SEGMENTOS

FDA centroides					**FDA *k*-segmentos *k=4***		
Classe	**Tamanho da classe**	**Classe predita**			**Classe predita**		
		1	**2**	**3**	**1**	**2**	**3**
1	59	**50**	0	9	**56**	3	0
		84,75%	0,00%	15,25%	**94,92%**	5,08%	0,00%
2	71	3	**49**	19	0	**70**	1
		4,23%	**69,01%**	26,76%	0,00%	**98,59%**	1,41%
3	48	0	17	**31**	0	0	**48**
		0,00%	35,42%	**64,58%**	0,00%	0,00%	**100,00%**
Casos classificados corretamente: 73,03%							**97,75%**

FONTE: O autor (2015)

A figura 58 mostra o gráfico das curvas principais para o conjunto wine alinhada com os escores do conjunto, o que explica o melhor desempenho para o algoritmo *k*-segmentos.

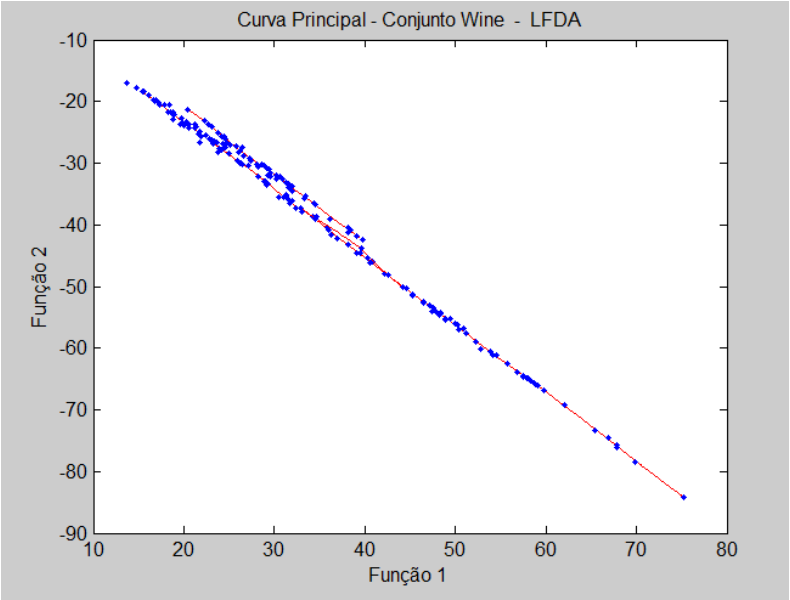


FIGURA 58 - CURVA PRINCIPAL PARA O CONJUNTO WINE - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.3.14 Veículo

Na análise da matriz de confusão (TABELA 31), o método por centroides apresentou resultado muito ruim, enquanto que o algoritmo *k*-segmento apresentou resultado inferior ao obtido pela FDA, porém próximo, apresentado na tabela 17 (página 98).

TABELA 30 - MATRIZ DE CONFUSÃO PARA O CONJUNTO VEÍCULO PARA FDA CENTROIDE X LFDA K-SEGMENTOS

FDA centroides						**FDA *k*-segmentos - *k* = 5**			
		Classe predita				**Classe predita**			
Classe	**Tamanho da classe**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1	218	**25**	131	9	53	**202**	2	8	6
		11,5%	60,1%	4,1%	24,3%	**92,7%**	0,9%	3,7%	2,7%
2	199	0	**199**	0	0	5	**178**	5	11

Continua

									Conclusão
		0,0%	**100,0%**	0,0%	0,0%	2,5%	**89,4%**	2,5%	5,5%
3	217	30	62	**28**	97	20	1	**134**	62
		13,8%	28,6%	**12,9%**	44,7%	9,2%	0,5%	**61,7%**	28,5%
4	212	25	60	22	**105**	18	2	65	**127**
		11,8%	28,3%	10,4%	**49,5%**	8,5%	0,9%	30,7%	**59,9%**
% de acertos:	**42,20%**					**% de acertos:**		**75,77%**	

FONTE: O autor (2015)

O gráfico das curvas principais apresentam a CP ajustada aos escores, a concentração dos escorres afetou a eficência na classificação (FIGURA 59).

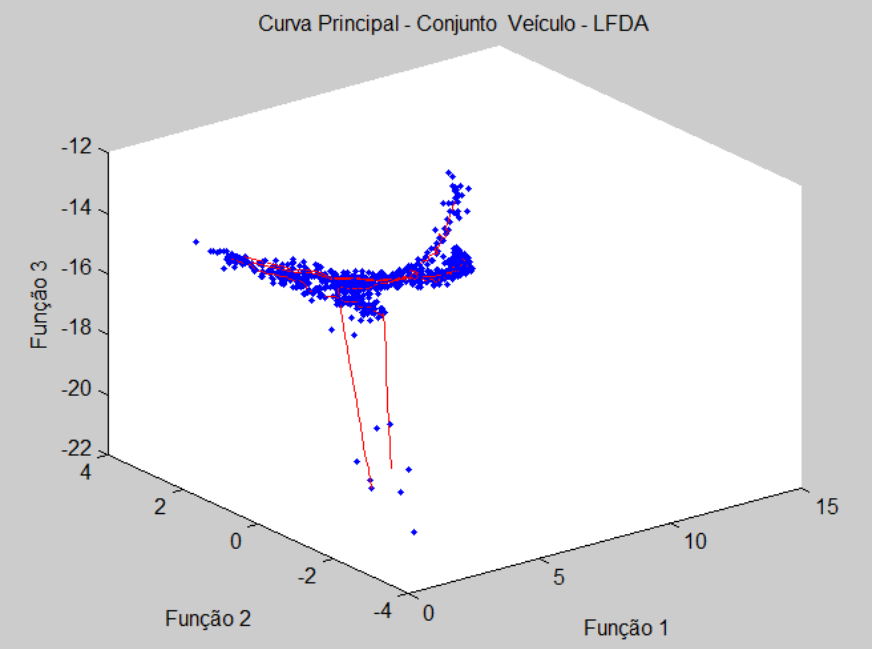


FIGURA 59 - CURVA PRINCIPAL PARA O CONJUNTO VEÍCULO - LFDA K-SEGMENTOS
FONTE: O autor (2015)

4.4 COMPARAÇÃO ENTRE LFDA E FDA PELO ALGORITMO *K*-SEGMENTO

4.4.1 Introdução

Nesta subseção é feita a comparação quanto à eficiência das técnicas multivariadas FDA e LFDA e também o desempenho das mesmas quando é utilizado o algoritmo *k*-segmentos.

O número de segmentos para a construção da curvas principais é considerado, com o objetivo de verificar a sua relevância quanto a eficiência da técnica *k*-segmentos. Em seguida é verificado o desempenho de cada técnica utilizada nos 14 conjuntos trabalhados.

4.4.2 Número de segmentos

A eficiência do algoritmo *k*-segmentos na classificação depende da quantidade de segmentos que formam a linha poligonal de cada classe e a CP. A quantidade de segmentos foi escolhida de forma empírica, de acordo com o seguinte critério: Para os conjuntos com número de observações pequeno foi estabelecido que o número de segmentos é 2 (mais um segmento de ligação). Para conjuntos com número maior de observações, determinaram-se 4 (mais 3 segmentos de ligação) segmentos por classe. A seguir, a tabela 32 apresenta os resultados para quantidades diferentes de segmentos por classe. É fácil ver que o número de segmentos por classe interfere no resultado final de classificação de cada técnica. Por exemplo, para o conjunto tiroide, quando é aplicada a técnica *k*-segmentos para FDA, se *k*=2, a eficiência é de 97,2%, para =3 é de 95,3% e se *k*=4 é de 97,2%. Para LFDA *k*-segmentos os resultados são: 94,8%, 95,3 e 96,7%. Para as técnicas com centroides são: 94,4% para FDA e 93,5% para LFDA. Da mesma forma pode-se verificar em análise para os demais conjuntos que o número de segmentos utilizados para a construção da linha poligonal é relevante quanto à eficiência do algoritmo *k*-segmentos, tanto para FDA, como para LFDA.

TABELA 31 - ALGORITMO K-SEGMENTOS APLICADO PARA QUANTIDADES DIFERENTES DE SEGMENTOS

conjunto	centroides		FDA k-segmentos			LFDA k-segmentos		
	FDA	LFDA	% (número de segmentos)			% (número de segmentos)		
1.Wine	100%	73,0%	100%(2)	100%(3)	100%(4)	93,8%(2)	97,2%(3)	97,7%(4)
2.Tiroide	94,4%	93,5%	97,2%(2)	95,3%(3)	97,2%(4)	94,8%(2)	95,3%(3)	96,7%(4)
3.Iris	98,0%	92,6%	98,6%(2)	98,0%(3)	98,6%(4)	97,3%(2)	98,6%(3)	98,6%(4)
4.Álcool	80,5%	68,8%	81,8%(2)	87,0%(3)	89,6%(4)	84,4%(2)	84,4%(3)	84,4%(4)
5.Glass	64,9%	48,5%	62,1%(2)	63,5%(3)	72,4%(5)	63,1%(2)	65,8%(3)	75,7%(5)
6.Wave	86,5%	80,0%	82,5%(2)	83,7%(3)	84,3%(5)	83,7(2)	85,6%(3)	85,7%(5)
7.Medical	88,8%	90,5%	92,5%(2)	92,5%(3)	92,5%(4)	92,3%(2)	90,3%(3)	90,5%(4)
8.Abalone	54,3%	48,4%	47,9%(2)	45,9%(3)	45,1%(4)	50,6%(2)	47,0%(3)	46,7%(4)
9.Besouro	100%	100%	100%(2)	100%(3)	100%(4)	100%(2)	100%(3)	100%(4)
10.Futebol	73,3%	68,8%	66,6%(2)	72,2%(3)	76,6%(4)	63,3%(2)	68,8%(3)	67,7%(4)
11.segm.	91,7%	78,7%	91,7%(3)	92,5%(4)	92,2%(5)	75,4%(3)	78,6%(4)	82,6%(5)
12.Letter	70,4%	56,5%	78,1%(3)	81,7%(4)	83,2%(6)	70,6%(3)	80,4%(4)	86,2%(6)
13.Bal.	69,3%	68,6%	91,0%(2)	91,0%(3)	91,0%(4)	91,0%(2)	91,0%(3)	92,4%(4)
14. Veíc.	77,2%	42,2%	76,0%(2)	77,3%(4)	77,1%(5)	70,7%(3)	75,3(4)	75,7%(5)

FONTE: O autor (2015)

Na tabela 33 são apresentados os resultados finais para os 14 conjuntos. É fácil ver o bom desempenho para FDA por *k*-segmentos, com melhor eficiência em 9 conjuntos dos 14 trabalhados. Comparando o algoritmo *k*-segmentos com a técnica dos centroides, tanto para FDA, como para LFDA o algoritmo foi superior. Na FDA o algoritmo *k*-segmentos não foi superior apenas no conjunto abalone, devido a grande concentração dos escores discriminantes. Da mesma forma, na LFDA o algoritmo *k*-segmento foi superior aos centroides, com resultado inferior apenas no conjunto futebol devido a proximidade das curvas nas classes 2 e 3. Já técnica LFDA por centroides apresentou resultado inferior dentre os métodos utilizados.

TABELA 32 - CLASSIFICAÇÃO FINAL DAS 4 TÉCNICAS PARA OS 14 CONJUNTOS

Conjunto	**Classificação**			
	FDA		**LFDA**	
	Centroide	***k*-seg.**	**Centroide**	***k*-seg.**
1.Wine	1º	1 º	3 º	2 º
2.Tiroide	3 º	1 º	4 º	2 º
3.Iris	2 º	1 º	3 º	1 º
4.Álcool	3 º	1 º	4 º	2 º
5.Glass	3 º	2 º	4 º	1 º
6.Wave	1 º	3 º	4 º	2 º
7.Medical	3º	1 º	2º	1 º
8.Abalone	1 º	4 º	3 º	2 º
9.Besouro	1 º	1 º	1 º	1 º
10.Futebol	3 º	2 º	1 º	4º
11.Segment	2 º	1 º	4º	3º
12.*Letter*	3 º	2 º	4 º	1 º
13.*Balance*	3 º	1 º	4 º	2 º
***14.*Veículo**	2 º	1 º	4º	3º

FONTE: O autor (2015)

5 CONCLUSÃO

O estudo tratou o problema de classificação de dados amostrais para conjuntos onde as classes (ou grupos) são conhecidas *a priori*, cuja eficiência do modelo utilizado é relevante.

A contribuição deste trabalho consistiu no desenvolvimento de um algoritmo para modificar a forma com que as técnicas de classificação FDA e LFDA classificam um novo objeto. A modificação consiste na troca dos centroides pelas linhas poligonais geradas pelas curvas principais. Esta técnica foi denominada classificador *k*-segmentos.

Os resultados experimentais obtidos pela técnica *k*-segmentos mostraram-se eficientes para conjuntos com 3 ou mais classes. Experimentalmente, o algoritmo proposto foi aplicado a 14 conjuntos amostrais e mostrou bom desempenho tanto para FDA como para LFDA, visto que a aplicação dessa técnica nos conjuntos apresentou melhor eficiência de classificação de novas observações amostrais. Comparando a FDA por centroides com *k*-segmentos, apenas em 3 conjuntos o método por centroides apresentou resultado superior, já para LFDA somente 1 conjunto obteve resultado melhor que o de *k*-segmentos. Consequentemente, a técnica *k*-segmentos foi superior a técnica dos centroides, tanto para FDA, como para LFDA. Na análise geral, a FDA com *k*-segmentos apresentou melhor eficiência que as demais técnicas em 11 conjuntos, com menor probabilidade de classificação incorreta.

A vantagem desse algoritmo é que ele não utiliza apenas um ponto central, como os centroides. A linha poligonal gerada pelo algoritmo pode determinar um ajuste não linear aos elementos da classe, o que possibilita melhorar a eficiência de classificação. Outra vantagem é quando o conjunto de escorres discriminantes das classes está concentrado em linha ou alguma forma de ajuste não linear, a CP é ajustada para forma do conjunto com os escores próximos a linha poligonal. A principal desvantagem dessa técnica é computacional, pois o algoritmo *k*-segmentos utiliza vários outros algoritmos e o desenvolvimento de um *software* é de complexidade muito superior ao dos centroides. O algoritmo pode não ser tão eficiente nos conjuntos quando: a transformação efetuada pela análise discriminante

não separa efetivamente as classes, isto é, a interseção entre classes é grande; quando a distribuição dos escores discriminantes tem distribuição esférica e também quando há o cruzamento das curvas principais.

Para pesquisas futuras pode-se estudar o número adequado de segmentos para o algoritmo (WANG; LEE, 2006). Outra possibilidade é o uso de outros algoritmos para a construção de curvas principais como CP de Hastie e Stuetzle e a NLPCA de Kramer.

REFERÊNCIAS

ANARAKI, F. P.; HUGHES, S. M. Efficient recovery of principal components from compressive measurements with application to Gaussian mixture model estimation. **IEEE – International Conference on Acoustic, Speech and Signal Processing (ICASSP)**, p. 2332-2336, 2014.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification and regression trees**. California: Wadsworth International, 1984.

BENVENISTE, S. M. **Investigation into text classification with kernel based schemes.** Dissertação (Master of Science Electrical Engineering) – Naval Postgraduate School, Monterey, CA, 2010.

CHIANG, L. H.; RUSSELL, E. L.; BRAATZ, R. D. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. Elsevier Science: **Chemometrics and Intelligent Laboratory systems**, v. 50, p. 243-252, 2000.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. 2. ed. New York: John Wiley & sons, 2001.

DUGARD, P.; TODMAN, J.; STAINES, H. **Data setes for approaching multivariante analysis: a pratical introduction**. London: Routledge, 2009.

FÁVERO, L. P.; BELFIORE, P.; DA SILVA, F. L.; CHAN, B. L. Análise de dados: modelagem multivariada para tomada de decisões. Rio de Janeiro, Elsevier, 2009.

FERREIRA, D. F. **Análise multivariada**. 1. ed. Lavras, MG: UFLA, 2008.

FUKUNAGA, K. **Introduction to statistical pattern recognition**. Boston: Academic Press, Inc., 1990.

FISHER, R. A. The use of multiple measurements in axonomic problems. **Annals of Eugenics**, v. 7, p. 179-188, 1936.

FREY, P. W.; SLATE, D. J. Letter recognition using holland-style adaptive classifiers. **Machine Learning**, v. 6, 1991.

GE, M.; FAN, L. Learning optimal kernel for pattern classification. **Wseas – Transactions on Mathematics**, 5. ed.,v. 12, p. 2224-2880, 2013.

GHAURI, S.A.; QURESHI, I. M.; AZIZ, M. A.; CHEEMA, T. A. Classification of digital modulated signals using linear discriminant analysis on faded channel. **World Applied Sciences Journal**, v. 29(10), p. 1220-1227, 2014.

GUO, Q.; CHEN, B.; JIANG, F.; JI, X.; KUNG, S. efficient divide-and-conquer classification based on feature-space decomposition. Cornell University Library, 2015.

HAIR, J. J. F.; ANDERSON, R. E.; TAHAM, R. L.; BLACK, W. C. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005.

HAND, D. J. Classifier technology and the illusion of progress. **Statistical Science,** v. 21, p. 1-15, 2006.

HASTIE, T. **Principal curves and Surfaces**. Tese (Pós-doutorado), Stanford Linear Accelerator Center, Stanford University, California, 1984.

HASTIE, T.; STUETZLE, W. Principal curves. **JASA Journal. American. Statistic. assoc.**, v. 84, p. 502–516, 1989.

HASTIE, T.; TIBSHIRANI, R. Discriminant adaptive nearest neighbor classification. **IEEE – Transactions on Pattern Analysis and Machine Intelligence**, v. 18, n. 6, 1996.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning**. 2. ed. New York, NY: Springer Science e Business Media LLC, 2009.

HSIEH, W. W. **Machine learning methods in the environmental sciences – neural networks and kernels.** Cambridge, UK: Cambridge University Press, 2009.

JOHNSON, R.; WICHERN, A., D. **Applied multivariate statistical analysis**, 4. ed. Upper Saddle River: Prentice Hall, 1998.

KALLAS, M.; FRANCIS, C.; KANAAN, L.; MERHEB, D.; HONEINE, P.; AMOUD, H. Multi-class SVM classification combined with kernel PCA feature of ECG signals. In: 19º INTERNATIONAL CONFERENCE ON TELECOMMUNICATIONS, Ottawa, 2012.

KÉGL, B.; KRZYZAK, A.; LINDER, T.; KENNETH, Z. Learning and design of principal curves. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 3, p. 281-297, 2000.

KITANI, E. C. **Análises discriminantes lineares para modelagem e reconstrução de imagens de faces**. Dissertação (Mestrado em Engenharia Elétrica), FEI, São Bernardo do Campo: SP, 2007.

KRAMER, A. M. Nonlinear principal component analysis using autoassociative neural networks. **AIChE journal**, v. 37, n. 2, fev. 1991.

KRUGER, U.; ZHANG, J.; XIE, L. Developments and applications of nonlinear principal component analysis – a review. **Principal for Data Visualization and Dimension Reduction lecture Notes in Computational Science and Enginee,** v. 58, 2008, p.1-43, 2008.

LACHENBRUCH, P. A.; MICKEY, M. R. **Estimation of error rates in discriminant analysis**. Technometrics, v. 10, n. 1, p. 1-11, 1968.

LAST, M.; TASSA, T.; ZHMUDYAK, A.; SHMUELI, E. Improving accuracy of classification models induced from anonymized datasets. **Elsevier Science Inc.** 256, p. 138-161, 2014.

LICCIARDI, G.; MARPU, P. R.; CHANUSSOT, J.; BENEDIKTSSON, J. A. Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. **IEEE-Geoscience and Remote Sensing Letters**, v. 9, n. 3, 2012.

LIM, T. S.; SHIH, Y. S. A Comparison of prediction accuracy, complexity, and training time of thirty-three. **Machine Learning** (Kluwer Academic Publishers, Boston), v. 40, p. 203-229, 2000.

LUBISCHEW, A. A. On the Use of Discriminant Functions in Taxonomy. **Biometrics**, v. 18, p. 455-477, 1962.

MATEUS, R. S.; MELO, R. O. L. de; FARIA, T. A. Análise de insolvência empresarial: uma abordagem financeira fundamentalista com aplicação do método estatístico multivariado e da técnica discriminante. **ReCont: Registro Contábil**. v. 2, n. 1, 2011.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: UFMG, 2005.

OKWONU, F. Z. Supervised learning techniques based on Fisher and Filter linear classification procedures for two groups problem. **International Journal of Mathematical Analysis and Applications**, v. 1(2), p. 27-30, 2014.

PANG, H.; TONG, T. Recent advances in discriminant analysis for high-dimensional data classification. **J. Biomet Biostat**, v. 3, 2012.

RAMLI, N. A.; ISMAIL, M. T.; WOOI, H. N. An analysis on two different data sets by using ensemble of k-nearest neighbor classifiers. **Wseas – Transactions on Mathematics**, v.13, p. 2224-2880, 2014.

RENCHER, A. C. **Methods of multivariate analysis**. 2. ed. New York: John Wiley & sons, 2002.

SHU, X.; LU, H. Linear discriminant analysis with spectral regularization. **Springer Science + business Media,** New York, p. 724-731, 2014.

SHLENS, J. A tutorial on principal components analysis. **Google Research**, Mountain View: CA, 2014.

SIEGLER, R. S. Three aspects of cognitive. Cognitive Psychology, v. 8, p. 481-520, 1976.

SUGIYAMA, M. Local Fisher discriminant analysis for supervised dimensionality reduction. **ICML06 Proceeding of the 23rdInternational Conference on Machine Learning**, p. 905-912, 2006.

SUGIYAMA, M. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. **Journal of Machine Learning Research**, v. 8, p. 1027-1061, 2007.

SUN, Z.; LI, J.; SUN, C. Kernel inverse Fisher discriminant analysis for face recognition. **Elsevier: Neurocomputing Letters**, v. 134, p. 46-52, 2014.

TANAGRA DATASETS. Disponível em:http:/eric.univlyon2.fr/~rico/Tanagra/ en/ tanagra.html. Acesso em 10/102014.

TARPEY, T.; FLURY, B. Self-consistency: a fundamental concept in statistics. **Statistics Science**, v. 11 n. 3, p. 229-243, 1996.

TIMM, N. H. **Applied multivariate analysis**. New York: Springer-Verlag New York, Inc., 2002.

UCI - Machines Learning Repository. Universidade da California de Irvine. Disponível em: <http://archive.ics.uci.edu/ml/index.html>. Acesso em: 10/04/2014.

VERBEEK, J. J.; VLASSIS, N.; KRÖSE, B. A k-segments algorithm for finding principal curves. **Elsevier: Pattern Recognition Letters**, v. 23, p. 1009–1017, 2002.

WANG, H.; LEE, T. C. M. Automatic parameter selection for a *k*-segment algorithm for computing principal curves. **Elsevier: Pattern Recognition Letters**, v. 27, p. 1142–1150, 2006.

WEBB, A. **Statisitical pattern recognition**. 2. ed. London: John Wiley e Sons, 2002.

WITTEN, D. M.; TIBSHIRANI, R. Penalized classification using Fisher's linear discriminant. **Journal of the Royal Statistical Society**, v. 73, part. 5, p. 753-772, 2011.

YUNSONG, L. P.; HUAIJIANG, Q. S. Microarrays data classification basead on principal curves. In: SEVENTH INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS AND KNOWLEDGE DISCOVERY, p. 2199-2202, 2010.

ZHAO, X.; LI, S. A modified kernel Fisher discriminant analysis algorithm for fault diagnosis. **International Journal of Advanced Computer Science,** v. 2, n. 1, p. 33-36, 2012.

APÊNDICES

APÊNDICE 1

TABELA - Matriz de confusão- Análise discriminante de Fisher para as variáveis 1 a 10 - Conjunto *Letter* - SPSS

Classe			Associação ao grupo prevista									
			1	2	3	4	5	6	7	8	9	10
Original	Contagem	1	677	1	0	0	0	0	0	11	0	16
		2	0	551	0	15	0	2	10	24	3	0
		3	0	5	548	0	26	9	57	9	0	0
		4	1	33	0	631	0	0	3	17	5	9
		5	0	49	97	2	342	7	84	1	4	0
		6	0	41	0	14	0	533	15	4	3	0
		7	9	38	128	1	16	0	382	15	0	1
		8	5	15	1	41	0	2	3	342	0	0
		9	0	15	1	10	5	9	2	0	617	12
		10	1	5	0	7	0	23	0	7	82	523
		11	0	15	9	10	30	0	15	6	1	0
		12	14	23	8	0	28	0	48	0	13	15
		13	9	3	0	0	0	0	0	13	0	2
		14	1	0	0	14	0	0	0	38	0	0
		15	15	3	2	58	0	1	12	95	0	2
		16	0	10	0	8	0	74	18	1	1	0
		17	19	40	0	3	5	0	51	8	0	7
		18	0	57	0	34	0	0	2	26	0	0
		19	21	76	0	6	2	22	26	1	3	10
		20	0	12	0	3	17	78	22	9	6	0
		21	0	0	4	2	0	0	0	47	0	0
		22	0	4	0	0	0	12	2	26		
		23	0	0	0	0	0	0	0	37	0	0
		24	0	18	0	14	7	0	3	2	7	1
		25	0	0	0	3	0	61	0	3	0	0
		26	2	1	0	0	35	6	7	1	1	21
	%	1	85,8	0,1	0	0	0	0	0	1,4	0	2
		2	0	71,9	0	2	0	0,3	1,3	3,1	0,4	0
		3	0	0,7	74,5	0	3,5	1,2	7,7	1,2	0	0
		4	0,1	4,1	0	78,4	0	0	0,4	2,1	0,6	1,1
		5	0	6,4	12,6	0,3	44,5	0,9	10,9	0,1	0,5	0
		6	0	5,3	0	1,8	0	68,8	1,9	0,5	0,4	0
		7	1,2	4,9	16,6	0,1	2,1	0	49,4	1,9	0	0,1
		8	0,7	2	0,1	5,6	0	0,3	0,4	46,6	0	0
		9	0	2	0,1	1,3	0,7	1,2	0,3	0	81,7	1,6

continua

Classe	Associação ao grupo prevista									
	1	2	3	4	5	6	7	8	9	10
										conclusão
10	0,1	0,7	0	0,9	0	3,1	0	0,9	11	70
11	0	2	1,2	1,4	4,1	0	2	0,8	0,1	0
12	1,8	3	1,1	0	3,7	0	6,3	0	1,7	2
13	1,1	0,4	0	0	0	0	0	1,6	0	0,3
14	0,1	0	0	1,8	0	0	0	4,9	0	0
15	2	0,4	0,3	7,7	0	0,1	1,6	12,6	0	0,3
16	0	1,2	0	1	0	9,2	2,2	0,1	0,1	0
17	2,4	5,1	0	0,4	0,6	0	6,5	1	0	0,9
18	0	7,5	0	4,5	0	0	0,3	3,4	0	0
19	2,8	10,2	0	0,8	0,3	2,9	3,5	0,1	0,4	1,3
20	0	1,5	0	0,4	2,1	9,8	2,8	1,1	0,8	0
21	0	0	0,5	0,2	0	0	0	5,8	0	0
22	0	0,5	0	0	0	1,6	0,3	3,4	0	0
23	0	0	0	0	0	0	0	4,9	0	0
24	0	2,3	0	1,8	0,9	0	0,4	0,3	0,9	0,1
25	0	0	0	0,4	0	7,8	0	0,4	0	0
26	0,3	0,1	0	0	4,8	0,8	1	0,1	0,1	2,9

a. 70,4% de casos originais agrupados corretamente classificados.

APÊNDICE 2

TABELA - Matriz de confusão- Análise discriminante de Fisher para as variáveis 11 a 20 - Conjunto *Letter* - SPSS

Classe			Associação ao grupo prevista									
			11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Original	Contagem	1	7	0	11	2	11	0	0	1	20	1
		2	5	0	1	0	17	0	0	72	43	0
		3	42	0	2	1	13	0	1	1	13	2
		4	1	0	7	11	24	0	0	21	15	0
		5	21	0	0	0	1	0	10	14	29	0
		6	1	0	0	3	1	84	15	2	29	8
		7	45	1	3	2	17	0	34	28	38	0
		8	54	0	3	61	62	7	21	45	0	0
		9	0	0	0	0	1	8	9	3	42	0
		10	1	0	0	1	13	12	16	0	40	0
		11	502	0	4	9	4	0	1	73	0	0
		12	12	558	1	0	2	0	6	2	15	0
		13	6	0	694	15	7	0	0	10	0	0
		14	11	0	16	639	7	0	0	2	0	1
		15	6	0	7	7	505	0	10	4	0	1
		16	6	0	1	1	12	605	20	1	3	3
		17	14	8	1	0	78	0	487	2	31	0
		18	38	0	2	5	5	0	0	568	0	0
		19	0	14	0	0	2	0	4	26	366	0
		20	14	0	0	0	9	7	0	2	18	567
		21	9	2	35	29	37	0	0	0	0	0
		22	11	0	2	0	9	2	0	3	0	8
		23	8	0	26	17	1	0	0	2	0	0
		24	10	0	0	0	2	1	32	8	54	2
		25	0	0	2	1	6	0	65	0	30	48
		26	0	0	0	0	0	0	20	3	101	1
	%	1	0,9	0	1,4	0,3	1,4	0	0	0,1	2,5	0,1
		2	0,7	0	0,1	0	2,2	0	0	9,4	5,6	0
		3	5,7	0	0,3	0,1	1,8	0	0,1	0,1	1,8	0,3
		4	0,1	0	0,9	1,4	3	0	0	2,6	1,9	0
		5	2,7	0	0	0	0,1	0	1,3	1,8	3,8	0

continua

Classe	Associação ao grupo prevista									
	11	12	13	14	15	16	17	18	19	20
										conclusão
6	0,1	0	0	0,4	0,1	10,8	1,9	0,3	3,7	1
7	5,8	0,1	0,4	0,3	2,2	0	4,4	3,6	4,9	0
8	7,4	0	0,4	8,3	8,4	1	2,9	6,1	0	0
9	0	0	0	0	0,1	1,1	1,2	0,4	5,6	0
10	0,1	0	0	0,1	1,7	1,6	2,1	0	5,4	0
11	67,9	0	0,5	1,2	0,5	0	0,1	9,9	0	0
12	1,6	73,3	0,1	0	0,3	0	0,8	0,3	2	0
13	0,8	0	87,6	1,9	0,9	0	0	1,3	0	0
14	1,4	0	2	81,6	0,9	0	0	0,3	0	0,1
15	0,8	0	0,9	0,9	67,1	0	1,3	0,5	0	0,1
16	0,7	0	0,1	0,1	1,5	75,3	2,5	0,1	0,4	0,4
17	1,8	1	0,1	0	10	0	62,2	0,3	4	0
18	5	0	0,3	0,7	0,7	0	0	74,9	0	0
19	0	1,9	0	0	0,3	0	0,5	3,5	48,9	0
20	1,8	0	0	0	1,1	0,9	0	0,3	2,3	71,2
21	1,1	0,2	4,3	3,6	4,6	0	0	0	0	0
22	1,4	0	0,3	0	1,2	0,3	0	0,4	0	1
23	1,1	0	3,5	2,3	0,1	0	0	0,3	0	0
24	1,3	0	0	0	0,3	0,1	4,1	1	6,9	0,3
25	0	0	0,3	0,1	0,8	0	8,3	0	3,8	6,1
26	0	0	0	0	0	0	2,7	0,4	13,8	0,1

a. 70,4% de casos originais agrupados corretamente classificados.

APÊNDICE 3

TABELA - Matriz de confusão- Análise discriminante de Fisher para as variáveis 21 a 26 - Conjunto *Letter* - SPSS

Classe			Associação ao grupo prevista					
			21	22	23	24	25	26
Original	Contagem	1	3	1	5	11	11	0
		2	0	0	1	17	0	5
		3	4	1	2	0	0	0
		4	0	0	0	27	0	0
		5	6	0	0	86	0	15
		6	2	1	5	11	3	0
		7	0	0	7	7	0	1
		8	20	7	4	39	2	0
		9	0	0	0	9	11	1
		10	0	0	0	15	0	1
		11	23	5	3	29	0	0
		12	0	0	0	16	0	0
		13	3	0	30	0	0	0
		14	1	3	46	4	0	0
		15	0	0	22	3	0	0
		16	1	0	13	0	25	0
		17	0	2	8	12	2	5
		18	0	0	0	21	0	0
		19	0	3	0	53	6	107
		20	3	5	0	12	4	8
		21	637	0	8	2	1	0
		22	4	652	20	2	7	0
		23	2	2	657	0	0	0
		24	26	1	0	579	12	8
		25	5	151	0	4	407	0
		26	0	0	0	15	0	520
	%	1	0,4	0,1	0,6	1,4	1,4	0
		2	0	0	0,1	2,2	0	0,7
		3	0,5	0,1	0,3	0	0	0
		4	0	0	0	3,4	0	0
		5	0,8	0	0	11,2	0	2
		6	0,3	0,1	0,6	1,4	0,4	0
		7	0	0	0,9	0,9	0	0,1
		8	2,7	1	0,5	5,3	0,3	0
		9	0	0	0	1,2	1,5	0,1
		10	0	0	0	2	0	0,1

continua

Classe	Associação ao grupo prevista					
						conclusão
11	3,1	0,7	0,4	3,9	0	0
	0	0	0	2,1	0	0
13	0,4	0	3,8	0	0	0
14	0,1	0,4	5,9	0,5	0	0
15	0	0	2,9	0,4	0	0
16	0,1	0	1,6	0	3,1	0
17	0	0,3	1	1,5	0,3	0,6
18	0	0	0	2,8	0	0
19	0	0,4	0	7,1	0,8	14,3
20	0,4	0,6	0	1,5	0,5	1
21	78,4	0	1	0,2	0,1	0
22	0,5	85,3	2,6	0,3	0,9	0
23	0,3	0,3	87,4	0	0	0
24	3,3	0,1	0	73,6	1,5	1
25	0,6	19,2	0	0,5	51,8	0
26	0	0	0	2	0	70,8

a. 70,4% de casos originais agrupados corretamente classificados.

APÊNDICE 4

TABELA - Matriz de confusão- FDA - *k*-segmentos para as variáveis 1 a 9 - Conjunto *Letter*.

		Classe	Classe predita							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	789	770	1	0	0	0	0	0	1	0
		97,5%	0,13%	0,00%	0,00%	0,00%	0,00%	0,00%	0,13%	0,00%
2	766	1	627	0	17	1	2	2	27	11
		0,13%	81,8%	0,00%	2,22%	0,13%	0,26%	0,26%	3,52%	1,44%
3	736	2	0	578	0	62	4	42	2	0
		0,27%	0,00%	78,5%	0,00%	8,42%	0,54%	5,71%	0,27%	0,00%
4	805	6	16	0	724	0	0	2	10	1
		0,75%	1,99%	0,00%	89,9%	0,00%	0,00%	0,25%	1,24%	0,12%
5	768	0	45	13	3	604	2	20	7	2
		0,00%	5,86%	1,69%	0,39%	78,6%	0,26%	2,60%	0,91%	0,26%
6	775	2	11	0	6	23	607	5	19	8
		0,26%	1,42%	0,00%	0,77%	2,97%	78,3%	0,65%	2,45%	1,03%
7	773	1	21	91	5	37	1	482	6	1
		0,13%	2,72%	11,7%	0,65%	4,79%	0,13%	62,3%	0,78%	0,13%
8	734	1	13	3	46	1	0	2	469	0
		0,14%	1,77%	0,41%	6,27%	0,14%	0,00%	0,27%	63,9%	0,00%
9	755	1	4	0	7	2	13	0	3	674
		0,13%	0,53%	0,00%	0,93%	0,26%	1,72%	0,00%	0,40%	89,2%
10	747	9	0	0	6	1	2	0	5	19
		1,20%	0,00%	0,00%	0,80%	0,13%	0,27%	0,00%	0,67%	2,54%
11	739	0	13	6	9	11	0	3	45	0
		0,00%	1,76%	0,81%	1,22%	1,49%	0,00%	0,41%	6,09%	0,00%
12	761	2	5	3	0	7	0	14	5	2
		0,26%	0,66%	0,39%	0,00%	0,92%	0,00%	1,84%	0,66%	0,26%
13	792	4	3	0	1	0	0	3	7	0
		0,51%	0,38%	0,00%	0,13%	0,00%	0,00%	0,38%	0,88%	0,00%
14	783	3	0	0	24	0	0	0	21	0
		0,38%	0,00%	0,00%	3,07%	0,00%	0,00%	0,00%	2,68%	0,00%
15	753	1	2	10	30	0	1	2	42	0
		0,13%	0,27%	1,33%	3,98%	0,00%	0,13%	0,27%	5,58%	0,00%
16	803	0	5	0	2	7	60	7	2	7

continua

		Classe	Classe predita							
		1	2	3	4	5	6	7	8	9
										conclusão
		0,00%	0,62%	0,00%	0,25%	0,87%	7,47%	0,87%	0,25%	0,87%
17	783	9	5	0	1	17	0	9	5	3
		1,15%	0,64%	0,00%	0,13%	2,17%	0,00%	1,15%	0,64%	0,38%
18	758	2	64	0	23	0	0	8	12	5
		0,26%	8,44%	0,00%	3,03%	0,00%	0,00%	1,06%	1,58%	0,66%
19	748	4	25	0	3	15	11	3	1	17
		0,53%	3,34%	0,00%	0,40%	2,01%	1,47%	0,40%	0,13%	2,27%
20	796	0	3	2	3	4	53	5	7	4
		0,00%	0,38%	0,25%	0,38%	0,50%	6,66%	0,63%	0,88%	0,50%
21	813	28	0	3	0	0	0	1	9	0
		3,44%	0,00%	0,37%	0,00%	0,00%	0,00%	0,12%	1,11%	0,00%
22	764	1	2	0	0	1	0	4	3	0
		0,13%	0,26%	0,00%	0,00%	0,13%	0,00%	0,52%	0,39%	0,00%
23	752	2	0	0	0	0	0	2	5	0
		0,27%	0,00%	0,00%	0,00%	0,00%	0,00%	0,27%	0,66%	0,00%
24	787	0	6	0	8	6	0	0	4	36
		0,00%	0,76%	0,00%	1,02%	0,76%	0,00%	0,00%	0,51%	4,57%
25	786	4	0	0	0	0	13	0	3	0
		0,51%	0,00%	0,00%	0,00%	0,00%	1,65%	0,00%	0,38%	0,00%
26	734	17	5	0	2	31	0	1	1	0
		2,32%	0,68%	0,00%	0,27%	4,22%	0,00%	0,14%	0,14%	0,00%
			% de acerto		83,2%					

APÊNDICE 5

TABELA - Matriz de confusão- FDA - *k*-segmentos para as variáveis 10 a 18 - Conjunto *Letter*.

		Classe	Classe predita							
		10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	2	0	3	3	0	0	0	0	0	0
	0,2%	0,00%	0,38%	0,38%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2	0	0	0	0	0	0	1	0	27	11
	0,0%	0,00%	0,00%	0,00%	0,00%	0,00%	0,13%	0,00%	3,52%	1,44%
3	0	4	11	0	0	11	0	1	2	0
	0,0%	0,54%	1,49%	0,00%	0,00%	1,49%	0,00%	0,14%	0,27%	0,00%
4	3	0	2	0	1	4	1	1	15	1
	0,3%	0,00%	0,25%	0,00%	0,12%	0,50%	0,12%	0,12%	1,86%	0,12%
5	0	0	25	0	0	0	0	1	3	2
	0,0%	0,00%	3,26%	0,00%	0,00%	0,00%	0,00%	0,13%	0,39%	0,26%
6	0	0	0	0	1	1	39	1	3	8
	0,0%	0,00%	0,00%	0,00%	0,13%	0,13%	5,03%	0,13%	0,39%	1,03%
7	0	14	18	2	0	14	2	30	10	1
	0,0%	1,81%	2,33%	0,26%	0,00%	1,81%	0,26%	3,88%	1,29%	0,13%
8	2	46	1	11	3	32	4	9	32	0
	0,2%	6,27%	0,14%	1,50%	0,41%	4,36%	0,54%	1,23%	4,36%	0,00%
9	14	0	6	0	1	0	4	0	0	674
	1,8%	0,00%	0,79%	0,00%	0,13%	0,00%	0,53%	0,00%	0,00%	89,2%
10	666	0	0	0	1	0	0	4	1	19
	89%	0,00%	0,00%	0,00%	0,13%	0,00%	0,00%	0,54%	0,13%	2,5%
11	0	539	10	0	0	3	0	1	31	0
	0,0%	72,9%	1,35%	0,00%	0,00%	0,41%	0,00%	0,14%	4,19%	0,00%
12	0	2	687	1	0	0	0	7	1	2
	0,0%	0,26%	90,2%	0,13%	0,00%	0,00%	0,00%	0,92%	0,13%	0,26%
13	0	10	0	714	8	10	0	0	3	0
	0,0%	1,26%	0,00%	90,1%	1,01%	1,26%	0,00%	0,00%	0,38%	0,00%
14	0	1	0	9	688	4	0	0	1	0
	0,0%	0,13%	0,00%	1,15%	87,8%	0,51%	0,00%	0,00%	0,13%	0,00%
15	2	0	0	1	0	604	9	25	0	0
	0,2%	0,00%	0,00%	0,13%	0,00%	80,2%	1,20%	3,32%	0,00%	0,00%
16	1	1	1	0	0	0	698	4	0	7
	0,1%	0,12%	0,12%	0,00%	0,00%	0,00%	86,9%	0,50%	0,00%	0,87%

continua

		Classe						Classe predita		
		10	11	12	13	14	15	16	17	18
										conclusão
17	9	0	8	1	0	63	6	621	2	3
	1,1%	0,00%	1,02%	0,13%	0,00%	8,05%	0,77%	79,3%	0,26%	0,38%
18	0	17	11	3	11	5	0	5	566	5
	0,0%	2,24%	1,45%	0,40%	1,45%	0,66%	0,00%	0,66%	74,6%	0,66%
19	5	0	14	0	0	0	1	4	4	17
	0,6%	0,00%	1,87%	0,00%	0,00%	0,00%	0,13%	0,53%	0,53%	2,27%
20	0	4	3	0	0	0	9	0	4	4
	0,0%	0,50%	0,38%	0,00%	0,00%	0,00%	1,13%	0,00%	0,50%	0,50%
21	0	1	0	14	2	6	0	0	0	0
	0,0%	0,12%	0,00%	1,72%	0,25%	0,74%	0,00%	0,00%	0,00%	0,00%
22	0	0	0	0	1	1	1	0	1	0
	0,0%	0,00%	0,00%	0,00%	0,13%	0,13%	0,13%	0,00%	0,13%	0,00%
23	0	0	0	9	2	11	0	0	0	0
	0,0%	0,00%	0,00%	1,20%	0,27%	1,46%	0,00%	0,00%	0,00%	0,00%
24	0	10	7	0	0	0	0	5	6	36
	0,0%	1,27%	0,89%	0,00%	0,00%	0,00%	0,00%	0,64%	0,76%	4,57%
25	0	0	0	1	0	0	41	2	0	0
	0,0%	0,00%	0,00%	0,13%	0,00%	0,00%	5,22%	0,25%	0,00%	0,00%
26	2	0	9	0	0	0	0	2	0	0
	0,2%	0,00%	1,23%	0,00%	0,00%	0,00%	0,00%	0,27%	0,00%	0,00%
			% de acerto		83,2%					

APÊNDICE 6

TABELA Matriz de confusão- FDA - *k*-segmentos para as variáveis 19 a 26 - Conjunto *Letter*.

Classe		**Classe Predita**							
K=6		**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**
1	0	0	0	2	3	0	4	0	0
	0,00%	0,00%	0,00%	0,25%	0,38%	0,00%	0,51%	0,00%	0,00%
2	11	0	2	35	0	2	0	0	11
	1,4%	0,00%	0,26%	4,57%	0,00%	0,26%	0,00%	0,00%	1,44%
3	0	3	6	3	5	0	0	0	0
	0,0%	0,41%	0,82%	0,41%	0,68%	0,00%	0,00%	0,00%	0,00%
4	3	0	2	1	1	12	0	0	3
	0,3%	0,00%	0,25%	0,12%	0,12%	1,49%	0,00%	0,00%	0,37%
5	7	8	0	0	0	22	0	6	7
	0,9%	1,04%	0,00%	0,00%	0,00%	2,86%	0,00%	0,78%	0,91%
6	5	26	1	4	2	2	8	1	5
	0,65%	3,35%	0,13%	0,52%	0,26%	0,26%	1,03%	0,13%	0,65%
7	10	1	2	13	9	3	0	0	10
	1,2%	0,13%	0,26%	1,68%	1,16%	0,39%	0,00%	0,00%	1,29%
8	1	3	13	8	0	27	2	5	1
	0,1%	0,41%	1,77%	1,09%	0,00%	3,68%	0,27%	0,68%	0,14%
9	10	1	0	0	0	6	0	9	10
	1,3%	0,13%	0,00%	0,00%	0,00%	0,79%	0,00%	1,19%	1,32%
10	10	0	0	0	0	11	0	12	10
	1,3%	0,00%	0,00%	0,00%	0,00%	1,47%	0,00%	1,61%	1,34%
11	6	1	9	5	3	44	0	0	6
	0,8%	0,14%	1,22%	0,68%	0,41%	5,95%	0,00%	0,00%	0,81%
12	5	0	0	1	0	19	0	0	5
	0,6%	0,00%	0,00%	0,13%	0,00%	2,50%	0,00%	0,00%	0,66%
13	0	0	4	8	17	0	0	0	0
	0,0%	0,00%	0,51%	1,01%	2,15%	0,00%	0,00%	0,00%	0,00%
14	0	0	0	26	6	0	0	0	0
	0,0%	0,00%	0,00%	3,32%	0,77%	0,00%	0,00%	0,00%	0,00%
15	0	1	12	0	11	0	0	0	0
	0,0%	0,13%	1,59%	0,00%	1,46%	0,00%	0,00%	0,00%	0,00%
16	3	0	0	2	1	2	0	0	3

continua

Classe		Classe Predita							
K=6		19	20	21	22	23	24	25	26
									conclusão
	0,3%	0,00%	0,00%	0,25%	0,12%	0,25%	0,00%	0,00%	0,37%
17	11	0	0	0	1	0	7	5	11
	1,4%	0,00%	0,00%	0,00%	0,13%	0,00%	0,89%	0,64%	1,40%
18	0	2	0	21	3	0	0	0	0
	0,0%	0,26%	0,00%	2,77%	0,40%	0,00%	0,00%	0,00%	0,00%
19	596	8	0	2	0	14	10	11	596
	79%	1,07%	0,00%	0,27%	0,00%	1,87%	1,34%	1,47%	79,6%
20	2	662	9	4	0	7	9	2	2
	0,2%	83,1%	1,13%	0,50%	0,00%	0,88%	1,13%	0,25%	0,25%
21	0	0	738	4	6	1	0	0	0
	0,0%	0,00%	90,7%	0,49%	0,74%	0,12%	0,00%	0	0,00%
22	0	1	2	730	13	0	3	0	0
	0,0%	0,13%	0,26%	95,5%	1,70%	0,00%	0,39%	0,00%	0,00%
23	0	0	9	14	698	0	0	0	0
	0,00%	0,00%	1,20%	1,86%	92,82%	0,00%	0,00%	0,00%	0,00%
24	9	5	6	0	0	674	3	2	9
	1,14%	0,64%	0,76%	0,00%	0,00%	85,64%	0,38%	0,25%	1,14%
25	3	26	4	47	1	4	637	0	3
	0,38%	3,31%	0,51%	5,98%	0,13%	0,51%	81,04%	0,00%	0,38%
26	62	7	0	0	0	10	0	585	62
	8,45%	0,95%	0,00%	0,00%	0,00%	1,36%	0,00%	79,70%	8,45%
			% de acerto		83,2%				

APÊNDICE 7

TABELA - Matriz de confusão- LFDA - centroides para as variáveis 1 a 9 - Conjunto *Letter*.

		Classe								Classe predita
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	789	**691**	22	0	0	0	0	1	4	0
		87,5 %	2,79 %	0,00 %	0,00 %	0,00 %	0,00%	0,13%	0,51%	0,00%
2	766	0	482	0	12	0	0	41	0	43
		0,00 %	62,9 %	0,00 %	1,57 %	0,00 %	0,00%	5,35%	0,00%	5,61%
3	736	0	9	531	2	19	23	19	6	4
		0,00 %	1,22 %	72,1 %	0,27 %	2,58 %	3,13%	2,58%	0,82%	0,54%
4	805	8	71	0	381	0	0	0	32	31
		0,99 %	8,82 %	0,00 %	47,3 %	0,00 %	0,00%	0,00%	3,98%	3,85%
5	768	0	46	4	0	340	16	37	0	68
		0,00 %	5,99 %	0,52 %	0,00 %	44,27 %	2,08%	4,82%	0,00%	8,85%
6	775	0	69	0	1	0	587	0	0	3
		0,00 %	8,90 %	0,00 %	0,13 %	0,00 %	75,7%	0,00%	0,00%	0,39%
7	773	15	23	139	14	12	1	394	11	0
		1,94 %	2,98 %	17,9 %	1,81 %	1,55 %	0,13%	50,9%	1,42%	0,00%
8	734	2	29	0	42	0	26	18	33	28
		0,27 %	3,95 %	0,00 %	5,72 %	0,00 %	3,54%	2,45%	4,50%	3,81%
9	755	0	14	0	4	5	13	0	0	549
		0,00 %	1,85 %	0,00 %	0,53 %	0,66 %	1,72%	0,00%	0,00%	72,7%
10	747	0	17	0	26	0	15	0	3	68
		0,00 %	2,28 %	0,00 %	3,48 %	0,00 %	2,01%	0,00%	0,40%	9,10%
11	739	2	37	12	2	52	2	36	17	37
		0,27 %	5,01 %	1,62 %	0,27 %	7,04 %	0,27%	4,87%	2,30%	5,01%
12	761	10	25	2	0	17	0	14	2	22
		1,31 %	3,29 %	0,26 %	0,00 %	2,23 %	0,0 0%	1,8 4%	0,2 6%	2,8 9%

continua

	Classe	Classe predita								
		1	2	3	4	5	6	7	8	9
										continuação
13	792	34	11	0	0	0	0	0	16	0
		4,29 %	1,39 %	0,00 %	0,00 %	0,00 %	0,00%	0,00%	2,02%	0,00%
14	783	16	7	0	22	0	14	0	50	1
		2,04 %	0,89 %	0,00 %	2,81 %	0,00 %	1,79%	0,00%	6,39%	0,13%
15	753	5	1	1	24	0	0	12	113	0
		0,66 %	0,13 %	0,13 %	3,19 %	0,00 %	0,00%	1,59%	15,01%	0,00%
16	803	0	23	0	3	0	154	3	23	10
		0,00 %	2,86 %	0,00 %	0,37 %	0,00 %	19,1%	0,37%	2,86%	1,25%
17	783	10	2	0	0	0	0	9	5	0
		1,28 %	0,26 %	0,00 %	0,00 %	0,00 %	0,00%	1,15%	0,64%	0,00%
18	758	0	48	0	47	0	0	11	5	20
		0,00 %	6,33 %	0,00 %	6,20 %	0,00 %	0,00%	1,45%	0,66%	2,64%
19	748	38	154	0	0	19	32	3	0	52
		5,08 %	20,5 %	0,00 %	0,00 %	2,54 %	4,28%	0,40%	0,00%	6,95%
20	796	0	22	0	1	22	99	22	0	1
		0,00 %	2,76 %	0,00 %	0,13 %	2,76 %	12,4%	2,76%	0,00%	0,13%
21	813	0	1	14	11	0	32	5	42	0
		0,00 %	0,12 %	1,72 %	1,35 %	0,00 %	3,94%	0,62%	5,17%	0,00%
22	764	0	23	0	1	0	0	0	20	0
		0,00 %	3,01 %	0,00 %	0,13 %	0,00 %	0,00%	0,00%	2,62%	0,00%
23	752	0	19	0	0	0	0	0	14	0
		0,00 %	2,53 %	0,00 %	0,00 %	0,00 %	0,00%	0,00%	1,86%	0,00%
24	787	0	7	0	2	109	10	12	2	85
		0,00 %	0,89 %	0,00 %	0,25 %	13,8 %	1,27%	1,52%	0,25%	10,8%
25	786	0	26	0	4	0	15	0	19	0
		0,00 %	3,31 %	0,00 %	0,51 %	0,00 %	1,91%	0,00%	2,42%	0,00%
26	734	2	35	0	0	87	3	0	0	0
		0,27 %	4,77 %	0,00 %	0,00 %	11,8 %	0,41%	0,00%	0,00%	0,00%

Classe	Classe predita								
	1	2	3	4	5	6	7	8	9
									conclusão
% de acerto								24,40%	

APÊNDICE 8

TABELA - Matriz de confusão- LFDA - centroides para as variáveis 10 a 18 - Conjunto *Letter*.

	Classe						**Classe predita**		
	10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	5	24	2	0	0	2	0	11	4
	0,63%	3,04%	0,25%	0,00%	0,00%	0,25%	0,00%	1,39%	0,51%
2	1	3	0	0	0	0	0	6	48
	0,13%	0,39%	0,00%	0,00%	0,00%	0,00%	0,00%	0,78%	6,27%
3	0	43	0	4	0	26	2	10	0
	0,00%	5,84%	0,00%	0,54%	0,00%	3,53%	0,27%	1,36%	0,00%
4	32	4	0	3	5	139	6	0	31
	3,98%	0,50%	0,00%	0,37%	0,62%	17,2%	0,75%	0,00%	3,85%
5	0	3	0	0	0	0	0	31	6
	0,00%	0,39%	0,00%	0,00%	0,00%	0,00%	0,00%	4,04%	0,78%
6	0	1	0	0	0	0	10	0	4
	0,00%	0,13%	0,00%	0,00%	0,00%	0,00%	1,29%	0,00%	0,52%
7	0	16	0	6	1	6	0	77	35
	0,00%	2,07%	0,00%	0,78%	0,13%	0,78%	0,00%	9,96%	4,53%
8	3	96	0	1	191	77	7	10	70
	0,41%	13,1%	0,00%	0,14%	26,0%	10,4%	0,95%	1,36%	9,54%
9	26	0	0	0	0	33	4	2	3
	3,44%	0,00%	0,00%	0,00%	0,00%	4,37%	0,53%	0,26%	0,40%
10	493	7	0	0	0	7	25	3	6
	66,0%	0,94%	0,00%	0,00%	0,00%	0,94%	3,35%	0,40%	0,80%
11	0	92	0	0	47	97	0	0	69
	0,00%	12,4%	0,00%	0,00%	6,36%	13,1%	0,00%	0,00%	9,34%
12	24	28	536	0	0	6	0	0	2
	3,15%	3,68%	70,43%	0,00%	0,00%	0,79%	0,00%	0,00%	0,26%
13	0	0	0	609	40	0	0	0	16
	0,00%	0,00%	0,00%	76,8%	5,05%	0,00%	0,00%	0,00%	2,02%
14	1	45	0	11	414	23	18	0	6
	0,13%	5,75%	0,00%	1,40%	52,8%	2,94%	2,30%	0,00%	0,77%
15	0	10	0	34	0	483	5	7	46
	0,00%	1,33%	0,00%	4,52%	0,00%	64,1%	0,66%	0,93%	6,11%
16	4	1	0	2	1	17	509	10	9

continua

	Classe	Classe predita							
	10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
									conclusão
	0,50%	0,12%	0,00%	0,25%	0,12%	2,12%	63,3%	1,25%	1,12%
17	0	0	10	7	0	128	0	486	23
	0,00%	0,00%	1,28%	0,89%	0,00%	16,3%	0,00%	62,0%	2,94%
18	18	58	0	0	0	73	0	12	436
	2,37%	7,65%	0,00%	0,00%	0,00%	9,63%	0,00%	1,58%	57,5%
19	20	1	6	0	0	0	1	1	3
	2,67%	0,13%	0,80%	0,00%	0,00%	0,00%	0,13%	0,13%	0,40%
20	0	28	0	0	0	0	0	1	0
	0,00%	3,52%	0,00%	0,00%	0,00%	0,00%	0,00%	0,13%	0,00%
21	0	44	1	29	61	34	0	6	4
	0,00%	5,41%	0,12%	3,57%	7,50%	4,18%	0,00%	0,74%	0,49%
22	0	0	0	12	2	0	3	0	13
	0,00%	0,00%	0,00%	1,57%	0,26%	0,00%	0,39%	0,00%	1,70%
23	0	0	0	54	12	0	0	0	16
	0,00%	0,00%	0,00%	7,18%	1,60%	0,00%	0,00%	0,00%	2,13%
24	1	98	0	0	0	12	0	49	0
	0,13%	12,4%	0,00%	0,00%	0,00%	1,52%	0,00%	6,23%	0,00%
25	0	0	0	5	0	2	21	64	0
	0,00%	0,00%	0,00%	0,64%	0,00%	0,25%	2,67%	8,14%	0,00%
26	9	0	0	0	0	0	1	0	3
	1,23%	0,00%	0,00%	0,00%	0,00%	0,00%	0,14%	0,00%	0,41%
		% de acerto		24,40%					

APÊNDICE 9

TABELA Matriz de confusão- LFDA - centroides para as variáveis 19 a 26 - Conjunto *Letter*.

Classe	Classe predita							
	19	**20**	**21**	**22**	**23**	**24**	**25**	**26**
1	1	0	0	0	8	2	12	0
	0,13%	0,00%	0,00%	0,00%	1,01%	0,25%	1,52%	0,00%
2	59	0	0	0	0	71	0	0
	7,70%	0,00%	0,00%	0,00%	0,00%	9,27%	0,00%	0,00%
3	32	5	0	0	0	1	0	0
	4,35%	0,68%	0,00%	0,00%	0,00%	0,14%	0,00%	0,00%
4	19	0	0	0	0	43	0	0
	2,36%	0,00%	0,00%	0,00%	0,00%	5,34%	0,00%	0,00%
5	61	1	0	0	0	128	0	27
	7,94%	0,13%	0,00%	0,00%	0,00%	16,67%	0,00%	3,52%
6	15	36	0	0	4	44	1	0
	1,94%	4,65%	0,00%	0,00%	0,52%	5,68%	0,13%	0,00%
7	22	0	0	0	1	0	0	0
	2,85%	0,00%	0,00%	0,00%	0,13%	0,00%	0,00%	0,00%
8	0	1	1	0	6	88	5	0
	0,00%	0,14%	0,14%	0,00%	0,82%	11,99%	0,68%	0,00%
9	31	0	0	0	0	71	0	0
	4,11%	0,00%	0,00%	0,00%	0,00%	9,40%	0,00%	0,00%
10	53	0	4	0	0	20	0	0
	7,10%	0,00%	0,54%	0,00%	0,00%	2,68%	0,00%	0,00%
11	0	41	1	1	1	180	13	0
	0,00%	5,55%	0,14%	0,14%	0,14%	24,36%	1,76%	0,00%
12	36	0	0	0	0	35	0	2
	4,73%	0,00%	0,00%	0,00%	0,00%	4,60%	0,00%	0,26%
13	0	0	11	1	54	0	0	0
	0,00%	0,00%	1,39%	0,13%	6,82%	0,00%	0,00%	0,00%
14	0	0	1	46	107	1	0	0
	0,00%	0,00%	0,13%	5,87%	13,67%	0,13%	0,00%	0,00%
15	0	0	6	0	6	0	0	0
	0,00%	0,00%	0,80%	0,00%	0,80%	0,00%	0,00%	0,00%
16	0	0	0	1	15	7	11	0

continuação

Classe	Classe predita							
	19	**20**	**21**	**22**	**23**	**24**	**25**	**26**
								conclusão
	0,00%	0,00%	0,00%	0,12%	1,87%	0,87%	1,37%	0,00%
17	77	0	0	3	7	1	5	10
	9,83%	0,00%	0,00%	0,38%	0,89%	0,13%	0,64%	1,28%
18	0	0	0	0	0	30	0	0
	0,00%	0,00%	0,00%	0,00%	0,00%	3,96%	0,00%	0,00%
19	227	1	0	0	0	106	0	84
	30,35%	0,13%	0,00%	0,00%	0,00%	14,17%	0,00%	11,23%
20	13	416	1	14	0	13	134	9
	1,63%	52,26%	0,13%	1,76%	0,00%	1,63%	16,83%	1,13%
21	0	32	480	4	4	0	9	0
	0,00%	3,94%	59,04%	0,49%	0,49%	0,00%	1,11%	0
22	0	0	0	654	30	0	6	0
	0,00%	0,00%	0,00%	85,60%	3,93%	0,00%	0,79%	0,00%
23	0	0	1	47	589	0	0	0
	0,00%	0,00%	0,13%	6,25%	78,32%	0,00%	0,00%	0,00%
24	88	62	0	0	0	249	0	1
	11,18%	7,88%	0,00%	0,00%	0,00%	31,64%	0,00%	0,13%
25	5	274	8	214	1	6	122	0
	0,64%	34,86%	1,02%	27,23%	0,13%	0,76%	15,52%	0,00%
26	79	0	0	0	0	8	0	507
	10,76%	0,00%	0,00%	0,00%	0,00%	1,09%	0,00%	69,07%
		% de acerto		24,40%				

APÊNDICE 10

TABELA - Matriz de confusão- LFDA - *k*-segmentos para as variáveis 1 a 9 - Conjunto *Letter*.

		Classe	Classe predita							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	789	**765**	0	0	0	0	1	0	0	1
		96,9%	0,00%	0,00%	0,00%	0,00%	0,13%	0,00%	0,00%	0,13%
2	766	2	**646**	0	12	7	6	3	9	2
		0,26%	**84,3%**	0,00%	1,57%	0,91%	0,78%	0,39%	1,17%	0,26%
3	736	0	0	**678**	0	5	1	17	1	0
		0,00%	0,00%	**92,1%**	0,00%	0,68%	0,14%	2,31%	0,14%	0,00%
4	805	6	19	0	**645**	0	4	0	29	1
		0,75%	2,36%	0,00%	**80,1%**	0,00%	0,50%	0,00%	3,60%	0,12%
5	768	1	6	14	0	**632**	19	29	2	1
		0,13%	0,78%	1,82%	0,00%	**82,2%**	2,47%	3,78%	0,26%	0,13%
6	775	0	1	0	4	4	**651**	2	5	22
		0,00%	0,13%	0,00%	0,52%	0,52%	**84,0%**	0,26%	0,65%	2,84%
7	773	5	5	72	4	7	1	**607**	7	0
		0,65%	0,65%	9,31%	0,52%	0,91%	0,13%	**78,5%**	0,91%	0,00%
8	734	3	10	7	34	0	6	5	**522**	0
		0,41%	1,36%	0,95%	4,63%	0,00%	0,82%	0,68%	**71,1%**	0,00%
9	755	0	1	0	0	5	3	0	0	**717**
		0,00%	0,13%	0,00%	0,00%	0,66%	0,40%	0,00%	0,00%	**94,9%**
10	747	4	0	0	7	2	1	0	2	45
		0,54%	0,00%	0,00%	0,94%	0,27%	0,13%	0,00%	0,27%	6,02%
11	739	0	1	19	8	11	3	5	24	0
		0,00%	0,14%	2,57%	1,08%	1,49%	0,41%	0,68%	3,25%	0,00%
12	761	0	1	1	0	3	0	6	0	2
		0,00%	0,13%	0,13%	0,00%	0,39%	0,00%	0,79%	0,00%	0,26%
13	792	2	2	0	3	0	0	13	2	0
		0,25%	0,25%	0,00%	0,38%	0,00%	0,00%	1,64%	0,25%	0,00%
14	783	8	2	0	26	0	0	0	6	0
		1,02%	0,26%	0,00%	3,32%	0,00%	0,00%	0,00%	0,77%	0,00%
15	753	1	0	12	16	0	0	1	10	0
		0,13%	0,00%	1,59%	2,12%	0,00%	0,00%	0,13%	1,33%	0,00%
16	803	0	3	0	2	0	78	4	8	8

continua

		Classe								Classe predita
		1	2	3	4	5	6	7	8	9
										conclusão
		0,00%	0,37%	0,00%	0,25%	0,00%	9,71%	0,50%	1,00%	1,00%
17	783	16	7	1	2	6	0	5	3	1
		2,04%	0,89%	0,13%	0,26%	0,77%	0,00%	0,64%	0,38%	0,13%
18	758	6	40	1	17	3	0	5	49	0
		0,79%	5,28%	0,13%	2,24%	0,40%	0,00%	0,66%	6,46%	0,00%
19	748	0	44	0	2	59	15	1	1	6
		0,00%	5,88%	0,00%	0,27%	7,89%	2,01%	0,13%	0,13%	0,80%
20	796	3	1	5	1	4	7	2	4	0
		0,38%	0,13%	0,63%	0,13%	0,50%	0,88%	0,25%	0,50%	0,00%
21	813	5	0	8	2	0	0	0	2	0
		0,62%	0,00%	0,98%	0,25%	0,00%	0,00%	0,00%	0,25%	0,00%
22	764	3	7	0	0	0	1	6	2	0
		0,39%	0,92%	0,00%	0,00%	0,00%	0,13%	0,79%	0,26%	0,00%
23	752	0	0	0	0	0	0	0	0	0
		0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
24	787	0	4	0	2	6	1	0	0	4
		0,00%	0,51%	0,00%	0,25%	0,76%	0,13%	0,00%	0,00%	0,51%
25	786	2	0	0	2	0	6	0	8	2
		0,25%	0,00%	0,00%	0,25%	0,00%	0,76%	0,00%	1,02%	0,25%
26	734	0	0	0	1	21	0	0	0	0
		0,00%	0,00%	0,00%	0,14%	2,86%	0,00%	0,00%	0,00%	0,00%
			% de acerto		86,24%					

APÊNDICE 11

TABELA - Matriz de confusão- LFDA - *k*-segmentos para as variáveis 10 a 18 - Conjunto *Letter*.

	Classe						**Classe predita**		
	10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	0	2	0	8	1	0	0	0	0
	0,00%	0,25%	0,00%	1,01%	0,13%	0,00%	0,00%	0,00%	0,00%
2	0	1	0	2	0	0	4	0	26
	0,00%	0,13%	0,00%	0,26%	0,00%	0,00%	0,52%	0,00%	3,39%
3	0	2	1	1	0	4	0	0	3
	0,00%	0,27%	0,14%	0,14%	0,00%	0,54%	0,00%	0,00%	0,41%
4	10	5	0	1	11	9	9	1	16
	1,24%	0,62%	0,00%	0,12%	1,37%	1,12%	1,12%	0,12%	1,99%
5	0	3	25	0	0	0	0	5	0
	0,00%	0,39%	3,26%	0,00%	0,00%	0,00%	0,00%	0,65%	0,00%
6	4	0	1	0	2	0	24	0	0
	0,52%	0,00%	0,13%	0,00%	0,26%	0,00%	3,10%	0,00%	0,00%
7	0	6	6	2	0	20	9	4	3
	0,00%	0,78%	0,78%	0,26%	0,00%	2,59%	1,16%	0,52%	0,39%
8	1	65	7	0	8	10	6	6	18
	0,14%	8,86%	0,95%	0,00%	1,09%	1,36%	0,82%	0,82%	2,45%
9	10	0	2	0	1	0	7	0	0
	1,32%	0,00%	0,26%	0,00%	0,13%	0,00%	0,93%	0,00%	0,00%
10	**661**	2	7	0	1	4	2	0	0
	88,4%	0,27%	0,94%	0,00%	0,13%	0,54%	0,27%	0,00%	0,00%
11	0	**581**	5	0	0	2	0	1	37
	0,00%	**78,6%**	0,68%	0,00%	0,00%	0,27%	0,00%	0,14%	5,01%
12	8	0	**724**	0	0	1	1	1	6
	1,05%	0,00%	**95,1%**	0,00%	0,00%	0,13%	0,13%	0,13%	0,79%
13	0	0	0	**722**	14	4	0	0	1
	0,00%	0,00%	0,00%	**91,1%**	1,77%	0,51%	0,00%	0,00%	0,13%
14	0	0	6	8	**687**	17	3	0	2
	0,00%	0,00%	0,77%	1,02%	**87,7%**	2,17%	0,38%	0,00%	0,26%
15	0	1	4	0	3	**656**	1	12	9
	0,00%	0,13%	0,53%	0,00%	0,40%	**87,1%**	0,13%	1,59%	1,20%
16	1	0	4	1	1	4	**669**	4	2

continua

	Classe	Classe predita							
	10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
									conclusão
	0,12%	0,00%	0,50%	0,12%	0,12%	0,50%	**83,3%**	0,50%	0,25%
17	0	0	7	0	0	54	2	**669**	0
	0,00%	0,00%	0,89%	0,00%	0,00%	6,90%	0,26%	**85,4%**	0,00%
18	0	13	17	1	16	1	2	7	**565**
	0,00%	1,72%	2,24%	0,13%	2,11%	0,13%	0,26%	0,92%	**74,5%**
19	2	3	3	0	0	3	2	1	1
	0,27%	0,40%	0,40%	0,00%	0,00%	0,40%	0,27%	0,13%	0,13%
20	0	4	4	0	0	0	3	1	4
	0,00%	0,50%	0,50%	0,00%	0,00%	0,00%	0,38%	0,13%	0,50%
21	0	0	0	4	16	7	0	0	0
	0,00%	0,00%	0,00%	0,49%	1,97%	0,86%	0,00%	0,00%	0,00%
22	0	0	0	3	1	1	0	0	1
	0,00%	0,00%	0,00%	0,39%	0,13%	0,13%	0,00%	0,00%	0,13%
23	0	0	0	5	6	6	1	3	0
	0,00%	0,00%	0,00%	0,66%	0,80%	0,80%	0,13%	0,40%	0,00%
24	3	20	27	0	0	0	1	3	0
	0,38%	2,54%	3,43%	0,00%	0,00%	0,00%	0,13%	0,38%	0,00%
25	4	0	0	0	0	1	15	4	0
	0,51%	0,00%	0,00%	0,00%	0,00%	0,13%	1,91%	0,51%	0,00%
26	1	0	3	0	0	0	0	9	0
	0,14%	0,00%	0,41%	0,00%	0,00%	0,00%	0,00%	1,23%	0,00%
	% de acerto							86,24%	

APÊNDICE 12

TABELA Matriz de confusão- LFDA - *k*-segmentos para as variáveis 19 a 26 - Conjunto *Letter*.

Classe	**Classe predita**							
	19	**20**	**21**	**22**	**23**	**24**	**25**	**26**
1	6	0	0	0	0	1	4	0
	0,76%	0,00%	0,00%	0,00%	0,00%	0,13%	0,51%	0,00%
2	14	1	1	26	0	4	0	0
	1,83%	0,13%	0,13%	3,39%	0,00%	0,52%	0,00%	0,00%
3	0	0	21	0	2	0	0	0
	0,00%	0,00%	2,85%	0,00%	0,27%	0,00%	0,00%	0,00%
4	21	0	4	4	0	6	0	4
	2,61%	0,00%	0,50%	0,50%	0,00%	0,75%	0,00%	0,50%
5	6	8	0	1	0	11	0	5
	0,78%	1,04%	0,00%	0,13%	0,00%	1,43%	0,00%	0,65%
6	4	43	0	0	0	1	7	0
	0,52%	5,55%	0,00%	0,00%	0,00%	0,13%	0,90%	0,00%
7	3	0	0	2	4	6	0	0
	0,39%	0,00%	0,00%	0,26%	0,52%	0,78%	0,00%	0,00%
8	0	1	11	10	3	0	0	1
	0,00%	0,14%	1,50%	1,36%	0,41%	0,00%	0,00%	0,14%
9	6	0	0	0	0	1	0	2
	0,79%	0,00%	0,00%	0,00%	0,00%	0,13%	0,00%	0,26%
10	4	0	0	0	0	0	0	5
	0,54%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,67%
11	5	0	5	1	1	29	0	1
	0,68%	0,00%	0,68%	0,14%	0,14%	3,92%	0,00%	0,14%
12	4	0	0	0	0	3	0	0
	0,53%	0,00%	0,00%	0,00%	0,00%	0,39%	0,00%	0,00%
13	0	0	10	3	16	0	0	0
	0,00%	0,00%	1,26%	0,38%	2,02%	0,00%	0,00%	0,00%
14	0	0	4	3	11	0	0	0
	0,00%	0,00%	0,51%	0,38%	1,40%	0,00%	0,00%	0,00%
15	0	2	16	0	9	0	0	0
	0,00%	0,27%	2,12%	0,00%	1,20%	0,00%	0,00%	0,00%
16	2	0	1	0	4	0	7	0

continua

Classe	Classe predita							
	19	**20**	**21**	**22**	**23**	**24**	**25**	**26**
								conclusão
	0,25%	0,00%	0,12%	0,00%	0,50%	0,00%	0,87%	0,00%
17	1	0	0	0	1	0	8	0
	0,13%	0,00%	0,00%	0,00%	0,13%	0,00%	1,02%	0,00%
18	0	0	1	8	3	3	0	0
	0,00%	0,00%	0,13%	1,06%	0,40%	0,40%	0,00%	0,00%
19	**587**	6	0	0	0	3	1	8
	78,48%	0,80%	0,00%	0,00%	0,00%	0,40%	0,13%	1,07%
20	1	**703**	4	5	0	8	31	1
	0,13%	**88,32%**	0,50%	0,63%	0,00%	1,01%	3,89%	0,13%
21	0	9	**749**	1	10	0	0	0
	0,00%	1,11%	**92,13%**	0,12%	1,23%	0,00%	0,00%	0
22	0	3	3	**715**	9	0	9	0
	0,00%	0,39%	0,39%	**93,59%**	1,18%	0,00%	1,18%	0,00%
23	0	0	2	5	**724**	0	0	0
	0,00%	0,00%	0,27%	0,66%	**96,28%**	0,00%	0,00%	0,00%
24	9	1	0	0	0	**704**	2	0
	1,14%	0,13%	0,00%	0,00%	0,00%	**89,45%**	0,25%	0,00%
25	3	73	3	33	8	1	**621**	0
	0,38%	9,29%	0,38%	4,20%	1,02%	0,13%	**79,01%**	0,00%
26	49	2	0	0	0	1	0	**647**
	6,68%	0,27%	0,00%	0,00%	0,00%	0,14%	0,00%	**88,15%**
	% de acerto			86,24%				

APÊNDICE 13

Programa para calcular as distâncias de um vetor a linha poligonal - FDA

```
%Dados necessarios:xdados:  Matriz de dados
%                  ks:      numero de segmentos
%                   Vclasse: Matriz 1 x g (numero de classes) com os
%tamanhos das classes
%saidas:           distanciaFisher:  distancia euclidiana do vetor a
%cada***
%                                    classe.
%                  classificaçãokseg:   distancia dos k segmentos ao
vetor.
%OBS: Para conjuntos com muitas classes pode ser necessario aumentar
o numero
%de segmentos na linha 46.


[Blda]=disc2(xdados,Vclasse);   %Calculo dos coeficientes da função
discriminante
[n,m]=size(xdados);
g=length(Vclasse);

for i=1:n
    d=distancia(xdados,Blda,  xdados(i,:),Vclasse);     %calculo  das
distancias do centróides (Fisher)
    distancias_Fisher(i,:)=d;
end
mc1=1;
mc2=Vclasse(1);

for i=1:g    %divide a matriz de dados em classes.
    ucl=xdados(mc1:mc2,:);
    eval(['classe' num2str(i) ' = ucl' ]);
    mc1=mc1+Vclasse(i);
    if i < g
        mc2=mc2+Vclasse(i+1);
    end
 end

for i=1:g
    mm=eval(['classe' num2str(i) ]);
    u=mm*Blda';                                    %Calculo do espaço
discriminante
    eval(['Ublda' num2str(i) '=u']); %Ublda para k-segments****
    [edges,vertices]=k_seg_soft(u,ks,1,1,1); %k-segmentos
    hold on

    eval(['edges' num2str(i) '=edges']);
    eval(['vertices' num2str(i) '=vertices']);
```

```
end
classificacaokseg=zeros(n,g);
mc2=Vclasse(1);
for cc=1:g
        mvertices=eval(['vertices'   num2str(cc)   ]);   %matriz   dos
vertices
        medges=eval(['edges' num2str(cc) ]);  % matriz dos nós
        for i=1:n
        vetor=xdados(i,:);
        vet=Blda*vetor';  %espaço discriminante da classe (grupo) 1

        %montagem dos vetores dos k_segmentos

        ne=length(medges);
        contador=1;

        for k=1:ne  % 2 arestas e 4 vertices matriz edges 4x4
            for j=1:ne
                 if k>j
                   if  medges(k,j)  ==2  |  medges(k,j)  ==1  %cria  o
vetor de %pontos dos k_segmentos

                        [d]=seg_dist(mvertices(:,j),  mvertices(:,k),
vet);
                        dd(contador,:)=d;
                        contador=contador+1;

                    end

                end
            end

        end

        xx=sort(dd);
        classificacaokseg(i,cc)=xx(1);
        end
    if cc< g
     mc2=Vclasse(cc+1)+ mc2;
    end
    end
    disp(classificacaokseg)
```

APÊNDICE 14

Programa para calcular as distâncias de um vetor a linha poligonal - LFDA

```
%Dados necessario: xdados:  Matriz de dados
%                  r:       numero de equações (escores) k-1
%                  ks:      numero de segmentos
%                  Vclasse: Matriz 1 x g com os tamanhos das classes
%saidas:           distancias_LFDA:  distancia euclidiana do vetor a
cada
%                                    classe.
%                     classificaçãoT:   distancia dos k segmentos ao
vetor.
%OBS: Para conjuntos com muitas classes pode ser necessario aumentar
o numero
%de segmentos na linha 46.
%*****************************************************************
*****
%
%***ClassificacaoT pelo algoritmo k-segmentos

m1=1;
m2=Vclasse(1);
g=length(Vclasse);
for j=1:g
    for i=m1:m2
        Y(i,1)=j;
    end
    m1=m2+1;
    if j < g
    m2=Vclasse(j+1)+m2;
    end
end
[T]=LFDA(xdados',Y,r, 'orthonormalized');
[n,m]=size(xdados);

for i=1:n
    d=distancia(xdados,T',xdados(i,:),Vclasse);
    distancias_LFDA(i,:)=d;

end
mc1=1;
mc2=Vclasse(1);

for i=1:g
    ucl=xdados(mc1:mc2,:);
    eval(['classe' num2str(i) ' = ucl' ]);
    mc1=mc1+Vclasse(i);
    if i < g
        mc2=mc2+Vclasse(i+1);
```

```
    end
end

for i=1:g
    mm=eval(['classe' num2str(i) ]);
    u=mm*T;                                          %Calculo do espaço
discriminante
    eval(['UbldaT' num2str(i) '=u']); %Ublda para k-segments****
    [edges,vertices]=k_seg_soft(u,ks,1,1,1); %k-segmentos
    hold on
    eval(['edgesT' num2str(i) '=edges']);
    eval(['verticesT' num2str(i) '=vertices']);
 end

classificacaoT=zeros(n,g);
mc2=Vclasse(1);

for cc=1:g
        mvertices=eval(['verticesT'  num2str(cc)  ]);  %matriz  dos
vertices
        medges=eval(['edgesT' num2str(cc) ]);  % matriz dos nós
        for i=1:n
            vetor=xdados(i,:);
            vet=T'*vetor';   %espaço discriminante da classe (grupo)
1

        %montagem dos vetores dos k_segmentos
        ne=length(medges);
        contador=1;

        for k=1:ne  % 2 arestas e 4 vertices matriz edges 4x4
            for j=1:ne
                if k>j
                    if  medges(k,j)  ==2  |  medges(k,j)  ==1  %cria  o
vetor de pontos dos k_segmentos
                       [d]=seg_dist(mvertices(:,j),   mvertices(:,k),
vet);
                        dd(contador,:)=d;
                        contador=contador+1;

                    end

                end
            end

        end

        xx=sort(dd);
        classificacaoT(i,cc)=xx(1);
        end
    if cc< g
     mc2=Vclasse(cc+1)+ mc2;
    end
    end
    disp(classificacaoT)
```